

From the INTERNATIONAL BUREAU

PCT  
NOTIFICATION OF ELECTION  
(PCT Rule 61.2)

To: Commissioner  
US Department of Commerce  
United States Patent and Trademark  
Office, PCT  
2011 South Clark Place Room  
CP2/5C24  
Arlington, VA 22202  
ETATS-UNIS D'AMERIQUE  
in its capacity as elected Office

Date of mailing (day/month/year)	30 May 2001 (30.05.01)
International application No.	PCT/US00/19948
International filing date (day/month/year)	21 July 2000 (21.07.00)
Priority date (day/month/year)	21 July 1999 (21.07.99)
Applicant's or agent's file reference	PF-0722 PCT
Applicant	HILLMAN, Jennifer, L. et al

1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on: 16 February 2001 (16.02.01)

☐ in a notice effecting later election filed with the International Bureau on: \_\_\_\_\_

2. The election ☒ was ☐ was not made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No.: (41-22) 740.14.35	Authorized officer Pascal Piriou Telephone No.: (41-22) 338.83.38
---	---



From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OR TRANSMITTAL  
OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

HAMLET-COX, Diana  
Incyte Genomics, Inc.  
3160 Porter Drive  
Palo Alto, CA 94304  
ETATS-UNIS D'AMERIQUE

To:

IMPORTANT NOTIFICATION

Date of mailing (day/month/year)  
23 April 2002 (23.04.02)

Applicant's or agent's file reference  
PF-0722 PCT

International application No.  
PCT/US00/19948

International filing date (day/month/year)  
21 July 2000 (21.07.00)

International publication date (day/month/year)  
01 February 2001 (01.02.01)

Priority date (day/month/year)  
21 July 1999 (21.07.99)

Applicant

INCYTE GENOMICS, INC. et al

1. The applicant is hereby notified of the date of receipt (except where the letters "NR" appear in the right-hand column) by the International Bureau of the priority document(s) relating to the earlier application(s) indicated below. Unless otherwise indicated by an asterisk appearing next to a date of receipt, or by the letters "NR", in the right-hand column, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).
2. This updates and replaces any previously issued notification concerning submission or transmission of priority documents.
3. An asterisk(\*) appearing next to a date of receipt, in the right-hand column, denotes a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b). In such a case, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.
4. The letters "NR" appearing in the right-hand column denote a priority document which was not received by the International Bureau or which the applicant did not request the receiving Office to prepare and transmit to the International Bureau, as provided by Rule 17.1(a) or (b), respectively. In such a case, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

Priority date	Priority application No.	Country or regional Office or PCT receiving Office	Date of receipt of priority document
21 July 1999 (21.07.99)	60/145,075	US	28 Marc 2002 (28.03.02)
08 Sept 1999 (08.09.99)	60/153,129	US	05 Apr 2002 (05.04.02)
10 Nov 1999 (10.11.99)	60/164,647	US	05 Apr 2002 (05.04.02)

The International Bureau of WIPO  
34, chemin des Colomбетtes  
1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer

Taieb AKREMI

Telephone No. (41-22) 338.83.38





INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference PF-0722 PCT	<b>FOR FURTHER ACTION</b> see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
International application No. PCT/US 00/19948	International filing date (day/month/year) 21/07/2000	(Earliest) Priority Date (day/month/year) 21/07/1999
Applicant INCYTE GENOMICS, INC. et al.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 6 sheets.  
☐ It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the report

- a. With regard to the language, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.
- ☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).
- b. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of the sequence listing :
- ☒ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☒ furnished subsequently to this Authority in computer readable form.
- ☒ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☒ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. ☒ Certain claims were found unsearchable (See Box I).

3. ☒ Unity of invention is lacking (see Box II).

4. With regard to the title,  
☒ the text is approved as submitted by the applicant.  
☐ the text has been established by this Authority to read as follows:

5. With regard to the abstract,  
☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the drawings to be published with the abstract is Figure No. \_\_\_\_\_  
☐ as suggested by the applicant.  
☐ because the applicant failed to suggest a figure.  
☐ because this figure better characterizes the invention.

☒ None of the figures.



# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/19948

A CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C12N5/10 C07K14/47 C07K16/18 C12Q1/68  
A61K38/17 G01N33/50 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 C12N C07K C12Q A61K G01N A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NEMOTO Y ET AL: "Recruitment of an alternatively spliced form of synaptotagmin 2 to mitochondria by the interaction with the PDZ domain of a mitochondrial outer membrane protein" EMBO JOURNAL, vol. 18, no. 11, 1 June 1999 (1999-06-01), pages 2991-3006, XP002156389 Rat OMP25: 88.966% identity in 145 aa overlap with SeqIdNo.1 / 75.835% identity in 1167 nt overlap with SeqIdNo.55 --- W0 98 45436 A (GENETICS INST) 15 October 1998 (1998-10-15) SeqIdNo.1414: 99.8% identity in 432 bp overlap with SeqIdNo.55 --- -/-	1,3,6,7, 9-11,13, 15,19, 22,25,26 3,11,12

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other documents, such combination being obvious to a person skilled in the art.
- "Z" document member of the same patent family

Date of the actual completion of the international search

4 January 2001

Date of mailing of the international search report

25.04.2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Lonnoy, O



# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/19948

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	Ep 1 033 401 A (GENSET) 6 September 2000 (2000-09-06) SeqIDNo.3623: 100.000% identity in 374 nt overlap with SeqIDNo.55 - & DATABASE GENSEQ [online] E.B.I., Hinxton, U.K.; Accession Number: C03625, 6 October 2000 (2000-10-06) DUMAS M ET AL: "Human secreted protein 5" EST, SeqIDNo.3623" XP002156390 abstract ----- MO 97 12962 A (COLD SPRING HARBOR LAB ; BEACH DAVID (US); CALIGIURI MAUREEN (US); ) 10 April 1997 (1997-04-10)	1,3,6,7, 9-15
T		
A		



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No  
PCT/US 00/19948

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
---	---------------------	----------------------------	---------------------

WO 9845436	A	AU 6891098 A	30-10-1998
		EP 0973896 A	26-01-2000
EP 1033401	A	NONE	
WO 9712962	A	US 6001619 A	14-12-1999
		EP 0857205 A	12-08-1998





INTERNATIONAL SEARCH REPORT

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see further information sheet invention group 1.

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.



(19) World Intellectual Property Organization  
International Bureau



(10) International Publication Number  
WO 01/07471 A2

(43) International Publication Date  
1 February 2001 (01.02.2001)

PCT

WO 01/07471 A2

(57) Abstract: The invention provides human cell cycle and proliferation proteins (CCYPR) and polynucleotides which identify and code CCYPR. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of CCYPR.

(54) Title: CELL CYCLE AND PROLIFERATION PROTEINS

- (51) International Patent Classification: C07K 14/00
- (21) International Application Number: PCT/US00/19948
- (22) International Filing Date: 21 July 2000 (21.07.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/145,075  
21 July 1999 (21.07.1999) US  
60/153,129  
8 September 1999 (08.09.1999) US  
60/164,647  
10 November 1999 (10.11.1999) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:  
60/145,075 (CIP)  
21 July 1999 (21.07.1999) US  
60/153,129 (CIP)  
8 September 1999 (08.09.1999) US  
60/164,647 (CIP)  
10 November 1999 (10.11.1999) US
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors; and  
(75) Inventors/Applicants (for US only): HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). TAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). TANG, Y. Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). BANDMAN, Olga
- (81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CZ, CU, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published: — Without international search report and to be republished upon receipt of that report.
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
- (74) Agents: HAMLETT-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).



# CELL CYCLE AND PROLIFERATION PROTEINS

## TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of cell cycle and proliferation proteins and to the use of these sequences in the diagnosis, treatment, and prevention of immune, developmental, and cell proliferative disorders including cancer.

## BACKGROUND OF THE INVENTION

Cell division is the fundamental process by which all living things grow and reproduce. In unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms, while in multicellular species many rounds of cell division are required to replace cells lost by wear or by programmed cell death, and for cell differentiation to produce a new tissue or organ. Details of the cell division cycle may vary, but the basic process consists of three principal events. The first event, interphase, involves preparations for cell division, replication of the DNA, and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell cytoplasm. The sequence and timing of cell cycle transitions are under the control of the cell cycle regulation system which controls the process by positive or negative regulatory circuits at various check points. Mitosis marks the end of interphase and concludes with the onset of cytokinesis. There are four stages in mitosis, occurring in the following order: prophase, metaphase, anaphase and telophase. Prophase includes the formation of bi-polar mitotic spindles, composed of microtubules and associated proteins such as dynein, which originate from polar mitotic centers. During metaphase, the nuclear material condenses and develops kinetochore fibers which aid in its physical attachment to the mitotic spindles. The ensuing movement of the nuclear material to opposite poles along the mitotic spindles occurs during anaphase. Telophase includes the disappearance of the mitotic spindles and kinetochore fibers from the nuclear material. Mitosis depends on the interaction of numerous proteins. For example, mutation studies in the *Drosophila melanogaster zw10* gene show a disruption in chromosome segregation. *ZW10* protein appears to function at the kinetochore as a tension-sensing checkpoint during the onset of anaphase. *ZW10* appears to have a direct role in the recruitment of dynein to the kinetochore, and, dynein's involvement in the coordination of chromosome separation at the onset of anaphase and/or poleward movement (Start, D.A. et al. (1998) J. Cell Biol. 142:763-774).

Regulated progression of the cell cycle depends on the integration of growth control pathways with the basic cell cycle machinery. Cell cycle regulators have been identified by selecting for human and yeast cDNAs that block or activate cell cycle arrest signals in the yeast mating

.

.

.

.

pheromone pathway when they are overexpressed. Known regulators include human CPR (cell cycle progression restoration) genes, such as CPR8 and CPR2, and yeast CDC (cell division control) genes, including CDC91, that block the arrest signals. The CPR genes express a variety of proteins including cyclins, tumor suppressor binding proteins, chaperones, transcription factors, translation factors, and RNA-binding proteins (Edwards, M.C. et al. (1997) Genetics 147:1063-1076).

The human CDC protein, CDC23, is homologous to the *S. cerevisiae* protein CDC23 which functions in the transition from metaphase to anaphase as well as in the exit from mitosis (Zhao, N. et al. (1998) Genomics 53:184-190). The *C. elegans* gene *cullin-1* (*cull1*) is a negative regulator of the cell cycle. *cull1* regulates the G1 to S phase transition and *C. elegans* *cull1* mutants exhibit hyperplasia of all tissues through acceleration of this transition by overriding mitotic arrest. *cull1* is a member of a conserved gene family that spans *S. cerevisiae*, nematodes and humans (Kipreos, E.T. et al. (1996) Cell 85:929-839).

Several cell cycle transitions, including the entry and exit of a cell from mitosis, are dependent upon the activation and inhibition of cyclin-dependent kinases (CDks). The CDks are composed of a kinase subunit, Cdk, and an activating subunit, cyclin, in a complex that is subject to many levels of regulation. There appears to be a single Cdk in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* whereas mammals have a variety of specialized CDks. Cyclins act by binding to and activating cyclin-dependent protein kinases which then phosphorylate and activate selected proteins involved in the mitotic process. The Cdk-cyclin complex is both positively and negatively regulated by phosphorylation, and by targeted degradation involving molecules such as CDK4 and CDK53. In addition, CDks are further regulated by binding to inhibitors and other proteins such as Suc1 that modify their specificity or accessibility to regulators (Patra, D. and W.G. Dunphy (1996) Genes Dev. 10:1503-1515; and Mathias, N. et al. (1996) Mol. Cell Biol. 16:6634-6643).

### Reproduction

The male and female reproductive systems are complex and involve many aspects of growth and development. The anatomy and physiology of the male and female reproductive systems are reviewed in Guyton, A.C. ((1991) Textbook of Medical Physiology, W.B. Saunders Co., Philadelphia PA, pp.899-928).

The male reproductive system includes the process of spermatogenesis, in which the sperm are formed. Male reproductive functions are regulated by various hormones. The hormones exert their effects on accessory sexual organs, and are involved in cellular metabolism, growth, and other bodily functions.

Spermatogenesis begins at puberty as a result of stimulation by gonadotropic hormones released from the anterior pituitary. Immature sperm (spermatogonia) undergo several mitotic cell divisions before undergoing meiosis and full maturation. The testes secrete several male sex hormones. Testosterone, the most abundant, is essential for growth and division of the immature sperm, and for the masculine characteristics of the male body. Three other male sex hormones,

.

.

.

.





gonadotropin-releasing hormone (GnRH), luteinizing hormone (LH), and follicle-stimulating

hormone (FSH), control sexual function.

The uterus, ovaries, fallopian tubes, vagina, and breasts comprise the female reproductive system. The ovaries and uterus are the source of ova and the location of fetal development,

respectively. The fallopian tubes and vagina are accessory organs attached to the top and bottom of the uterus, respectively. Both the uterus and ovaries have additional roles in the development and

loss of reproductive capability during a female's lifetime. The primary role of the breasts is lactation. Multiple endocrine signals from the ovaries, uterus, pituitary, hypothalamus, adrenal glands, and other tissues coordinate reproduction and lactation. These signals vary during the monthly

menstruation cycle and during the female's lifetime. Similarly, the sensitivity of reproductive organs to these endocrine signals varies during the female's lifetime.

A combination of positive and negative feedback to the ovaries, pituitary and hypothalamus glands controls physiologic changes during the monthly ovulation and endometrial cycles. The

anterior pituitary secretes two major gonadotropin hormones, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), regulated by negative feedback of steroids, most notably by ovarian

estradiol. If fertilization does not occur, estrogen and progesterone levels decrease. This sudden reduction of the ovarian hormones leads to menstruation, the desquamation of the endometrium.

Hormones further govern all the steps of pregnancy, parturition, lactation, and menopause.

During pregnancy large quantities of human chorionic gonadotropin (hCG), estrogens, progesterone, and human chorionic somatomammotropin (hCS) are formed by the placenta. hCG, a glycoprotein

similar to luteinizing hormone, stimulates the corpus luteum to continue producing more progesterone and estrogens, rather than to involute as occurs if the ovum is not fertilized. hCS is similar to growth

hormone and is crucial for fetal nutrition.

The female breast also matures during pregnancy. Large amounts of estrogen secreted by the placenta trigger growth and branching of the breast milk ductal system while lactation is initiated by

the secretion of prolactin by the pituitary gland.

Parturition involves several hormonal changes that increase uterine contractility toward the

end of pregnancy, as follows. The levels of estrogens increase more than those of progesterone.

Oxytocin is secreted by the neurohypophysis. Concomitantly, uterine sensitivity to oxytocin

increases. The fetus itself secretes oxytocin, cortisol (from adrenal glands), and prostaglandins.

Menopause occurs when most of the ovarian follicles have degenerated. The ovary then

produces less estradiol, reducing the negative feedback on the pituitary and hypothalamus glands.

Mean levels of circulating FSH and LH increase, even as ovulatory cycles continue. Therefore, the

ovary is less responsive to gonadotropins, and there is an increase in the time between menstrual

cycles. Consequently, menstrual bleeding ceases, and reproductive capability ends.

## Differentiation and Proliferation

Tissue growth involves complex and ordered patterns of cell proliferation, cell



differentiation, and apoptosis. Cell proliferation must be regulated to maintain both the number of cells and their spatial organization. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation.

5 Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, tumor-suppressor proteins, and mitosis-promoting factors.

Embryogenesis is a process in which distinct patterns of protein expression control proper development. This process involves a host of proteins each with distinct and highly coordinated expression patterns. For example, in the mouse, temporally regulated expression of two related genes 10 *Msl1* and *Mrl1* contribute to normal embryonic development. *Msl1* is expressed in the posterior domains of the developing mesoderm, while *Mrl1* is expressed in the anterior visceral endoderm. Properly coordinated expression of each protein throughout embryogenesis is critical for proper tissue and organ formation (Dunwoodie, S.L. et al. (1998) *Mech. Dev.* 72:27-40).

15 Growth factors were originally described as serum factors required to promote cell proliferation. Most growth factors are large, secreted polypeptides that act on cells in their local environment. Growth factors bind to and activate specific cell surface receptors and initiate intracellular signal transduction cascades. Many growth factor receptors are classified as receptor tyrosine kinases which undergo autophosphorylation upon ligand binding. Autophosphorylation 20 enables the receptor to interact with signal transduction proteins characterized by the presence of SH2 or SH3 domains (Src homology regions 2 or 3). These proteins then modulate the activity state of small G-proteins, such as Ras, Rab, and Rho, along with GTPase activating proteins (GAPs), guanine nucleotide releasing proteins (GNRPs), and other guanine nucleotide exchange factors. Small G proteins act as molecular switches that activate other downstream events, such as mitogen-activated protein kinase (MAP kinase) cascades. MAP kinases ultimately activate transcription of mitosis- 25 promoting genes.

In addition to growth factors, small signaling peptides and hormones also influence cell proliferation. These molecules bind primarily to another class of receptor, the trimeric G-protein coupled receptor (GPCR), found predominantly on the surface of immune, neuronal and 30 neuroendocrine cells. Upon ligand binding, the GPCR activates a trimeric G protein which in turn triggers increased levels of intracellular second messengers such as phospholipase C,  $Ca^{2+}$ , and cyclic AMP. Most GPCR-mediated signaling pathways indirectly promote cell proliferation by causing the secretion or breakdown of other signaling molecules that have direct mitogenic effects. These signaling cascades often involve activation of kinases and phosphatases. Some growth factors, such 35 as some members of the transforming growth factor beta (TGF- $\beta$ ) family, act on some cells to stimulate cell proliferation and on other cells to inhibit it. Growth factors may also stimulate a cell at one concentration and inhibit the same cell at another concentration. Most growth factors also have a



multitude of other actions besides the regulation of cell growth and division: they can control the proliferation, survival, differentiation, migration, or function of cells depending on the circumstance. For example, the tumor necrosis factor/nerve growth factor (TNF/NGF) family can activate or inhibit cell death, as well as regulate proliferation and differentiation. The cell response depends on the type of cell, its stage of differentiation and transformation status, which surface receptors are stimulated, and the types of stimuli acting on the cell (Smith, A. et al. (1994) Cell 76:959-962; and Nocentini, G. et al. (1997) Proc. Natl. Acad. Sci. USA 94:6216-6221).

Neighboring cells in a tissue compete for growth factors, and when provided with "unlimited" quantities in a perfused system will grow to even higher cell densities before reaching density-dependent inhibition of cell division. Cells often demonstrate an anchorage dependence of cell division as well. This anchorage dependence may be associated with the formation of focal contacts linking the cytoskeleton with the extracellular matrix (ECM). The expression of ECM components can be stimulated by growth factors. For example, TGF- $\beta$  stimulates fibroblasts to produce a variety of ECM proteins, including fibronectin, collagen, and tenascin (Pearson, C.A. et al. (1988) EMBO J. 7:2977-2981). In fact, for some cell types, specific ECM molecules, such as laminin or fibronectin, may act as growth factors. Tenascin-C and -R, expressed in developing and lesioned neural tissue, provide stimulatory/anti-adhesive or inhibitory properties, respectively, for axonal growth (Faisssner, A. (1997) Cell Tissue Res. 290:331-341).

Cancers and immune disorders are characterized by uncoordinated cell proliferation. Cancers are associated with the activation of oncogenes which are derived from normal cellular genes. These oncogenes encode oncoproteins which convert normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein, and other oncoproteins are abnormally expressed with respect to location or amount of expression. The latter category of oncoprotein causes cancer by altering transcriptional control of cell proliferation. Five classes of oncoproteins are known to affect cell cycle controls. These classes include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Certain cell proliferation disorders can be identified by changes in the protein complexes that normally control progression through the cell cycle. A primary treatment strategy involves reestablishing control over cell cycle progression by manipulation of the proteins involved in cell cycle regulation (Nigg, E.A. (1995) BioEssays 17:471-480).

Many oncogenes have been identified and characterized. These include sis, erbA, erbB, her-2, mutated G<sub>s</sub>, src, abl, ras, crk, jun, fos, myc, and mutated tumor-suppressor genes such as RB, p53, mdm2, Cip1, p16, and cyclin D. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the



breakpoint cluster region (bcr) on chromosome 22.

Mutations which hyperactivate oncogenes result in cell proliferation. Stimulation of a cell by growth factors activates two sets of gene products, the early-response genes and the delayed-response genes. Early-response gene products include *myc*, *fos*, and *jun*, all of which encode gene regulatory proteins. These regulatory proteins lead to the transcriptional activation of a second set of genes, the delayed-response genes, which include the cell-cycle regulators Cdk and cyclins. For example, the human T-cell leukemia virus type I (HTLV-1) Tax transactivator protein acts as an early response gene by enhancing the activity of a cellular transcription factor. The oncogenic properties of the Tax protein include transformation of primary T-lymphocytes and fibroblasts through cooperation with the a GTP- binding protein, Ras. Recently investigators have shown that Tax interacts with several PDZ-containing proteins. The PDZ domain, originally described in the *Drosophila* tumor suppressor protein Discs-Large, is common to membrane proteins thought to be involved in clustering receptors in growth factor signal transduction pathways (Rousset, R. et al. (1998) *Oncogene* 16:643-654).

Tumor-suppressor genes are involved in regulating cell proliferation. Mutations which cause reduced or loss of function in tumor-suppressor genes result in uncontrolled cell proliferation. For example, the retinoblastoma gene product (RB), in a non-phosphorylated state, binds several early-response genes and suppresses their transcription, thus blocking cell division. Phosphorylation of RB causes it to dissociate from the genes, releasing the suppression, and allowing cell division to proceed.

Other gene products involved in cell proliferation, differentiation, and apoptosis are yet to be discovered. One method currently being utilized to help identify such new molecules involves comparisons between quiescent and proliferative tissues. For example, a subtractive hybridization screen of human placental cytotrophoblast cells identified 20 genes whose expression levels rose due to EGF induction of cell proliferation. (Morrish, D.W. et al. (1996) *Placenta* 17:431-441). Another method involves identification of molecules produced in cells treated with anti-tumorigenic agents, such as dithiolethiones. Presumably, the protective action of these anti-tumorigenic agents is associated with the induction of tumor suppressor gene products (Primiano, T. et al. (1996) *Carcinogenesis* 17:2297-2303).

In another example, the candidate tumor-suppressor gene INGI, that codes a nuclear protein, p33ING1, is involved in the negative regulation of cell proliferation. The action of p33ING1 is dependent upon the activity of another tumor-suppressor gene, p53. p53 is a cellular stress-responsive gene requiring the activity of p33ING1 to effectively induce growth inhibition of cells. p33ING1 and p53 have been shown to physically associate through immunoprecipitation studies (Garavito, I. et al. (1998) *Nature* 391:295-298).

## Apoptosis

Apoptosis is the genetically controlled process by which unneeded or defective cells undergo programmed cell death. Selective elimination of cells is as important for morphogenesis and tissue





remodeling as is cell proliferation and differentiation. Lack of apoptosis may result in hyperplasia and other disorders associated with increased cell proliferation. Apoptosis is also a critical component of the immune response. Immune cells such as cytotoxic T-cells and natural killer cells prevent the spread of disease by inducing apoptosis in tumor cells and virus-infected cells. In addition, immune cells that fail to distinguish self molecules from foreign molecules must be eliminated by apoptosis to avoid an autoimmune response.

Apoptotic cells undergo distinct morphological changes. Hallmarks of apoptosis include cell shrinkage, nuclear and cytoplasmic condensation, and alterations in plasma membrane topology. Biochemically, apoptotic cells are characterized by increased intracellular calcium concentration, fragmentation of chromosomal DNA, and expression of novel cell surface components.

The molecular mechanisms of apoptosis are highly conserved, and many of the key protein regulators and effectors of apoptosis have been identified. Apoptosis generally proceeds in response to a signal which is transduced intracellularly and results in altered patterns of gene expression and protein activity. Signaling molecules such as hormones and cytokines are known both to stimulate and to inhibit apoptosis through interactions with cell surface receptors. Transcription factors also play an important role in the onset of apoptosis. A number of downstream effector molecules, particularly proteases such as the cysteine proteases called caspases, have been implicated in the degradation of cellular components and the proteolytic activation of other apoptotic effectors.

**Aging and Senescence**

Studies of the aging process or senescence have shown a number of characteristic cellular and molecular changes (Fauci, A.S. et al. (1998) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, p.37). These characteristics include increases in chromosome structural abnormalities, DNA cross-linking, incidence of single-stranded breaks in DNA, losses in DNA methylation, and degradation of telomere regions. In addition to these DNA changes, post-translational alterations of proteins increase including deamidation, oxidation, cross-linking, and nonenzymatic glycosylation. Still further molecular changes occur in the mitochondria of aging cells through deterioration of structure. These changes eventually contribute to decreased function in every organ of the body.

The discovery of new cell cycle and proliferation proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer.

**SUMMARY OF THE INVENTION**

The invention features purified polypeptides, cell cycle and proliferation proteins, referred to collectively as "CCYPR" and individually as "CCYPR-1," "CCYPR-2," "CCYPR-3," "CCYPR-4," "CCYPR-5," "CCYPR-6," "CCYPR-7," "CCYPR-8," "CCYPR-9," "CCYPR-10," "CCYPR-11,"



WO 01/07471  
 PCT/US00/19948

18, "CCYPR-19," "CCYPR-20," "CCYPR-21," "CCYPR-22," "CCYPR-23," "CCYPR-24,"  
 "CCYPR-25," "CCYPR-26," "CCYPR-27," "CCYPR-28," "CCYPR-29," "CCYPR-30," "CCYPR-31," "CCYPR-32," "CCYPR-33," "CCYPR-34," "CCYPR-35," "CCYPR-36," "CCYPR-37,"  
 "CCYPR-38," "CCYPR-39," "CCYPR-40," "CCYPR-41," "CCYPR-42," "CCYPR-43," "CCYPR-44," "CCYPR-45," "CCYPR-46," "CCYPR-47," "CCYPR-48," "CCYPR-49," "CCYPR-50,"  
 "CCYPR-51," "CCYPR-52," "CCYPR-53," "CCYPR-54." In one aspect, the invention provides an  
 isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an  
 amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring  
 amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from  
 the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence  
 selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino  
 acid sequence selected from the group consisting of SEQ ID NO:1-54. In one alternative, the  
 invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-54.  
 15 The invention further provides an isolated polynucleotide encoding a polypeptide comprising  
 an amino acid sequence selected from the group consisting of a) an amino acid sequence selected  
 from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having  
 at least 90% sequence identity to an amino acid sequence selected from the group  
 ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group  
 consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected  
 from the group consisting of SEQ ID NO:1-54. In one alternative, the polynucleotide encodes a  
 polypeptide selected from the group consisting of SEQ ID NO:1-54. In another alternative, the  
 polynucleotide is selected from the group consisting of SEQ ID NO:55-108.

20 Additionally, the invention provides a recombinant polynucleotide comprising a promoter  
 sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid  
 sequence selected from the group consisting of a) an amino acid sequence selected from the group  
 consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90%  
 sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54,  
 c) a biologically active fragment of an amino acid sequence selected from the group consisting of  
 SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the  
 group consisting of SEQ ID NO:1-54. In one alternative, the invention provides a cell transformed  
 with the recombinant polynucleotide. In another alternative, the invention provides a transgenic  
 organism comprising the recombinant polynucleotide.

25 The invention also provides a method for producing a polypeptide comprising an amino acid  
 sequence selected from the group consisting of a) an amino acid sequence selected from the group  
 consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90%  
 sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54,

30

35



- c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.
- 5 Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54.
- 10 The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence complementary to a), d) a polynucleotide complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.
- 20 Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence having at least consisting of SEQ ID NO:55-108, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence complementary to a), d) a polynucleotide complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.
- 30 The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence having at least consisting of SEQ ID NO:55-108, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence complementary to a), d) a polynucleotide complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.
- 35 The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence having at least consisting of SEQ ID NO:55-108, b) a naturally occurring polynucleotide sequence having at least



70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a pharmaceutical composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and a pharmaceutically acceptable excipient. In one embodiment, the pharmaceutical composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CCYP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CCYP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino





acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CYP<sub>PR</sub>, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:55-108, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.



The invention further provides a method for assessing toxicity of a test compound, said

method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID

NO:55-108, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, iii) a

polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific

hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group

consisting of SEQ ID NO:55-108, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID

NO:55-108, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence

complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of the above polynucleotide sequence; c) quantifying the amount of

hybridization complex; and d) comparing the amount of hybridization complex in the treated

biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is

indicative of toxicity of the test compound.

## BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOS),

clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding CCYPR.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of CCYPR.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was

cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding CCYPR were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold

parameters.



## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

"CCYP<sup>R</sup>" refers to the amino acid sequences of substantially purified CCYP<sup>R</sup> obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CCYP<sup>R</sup>. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CCYP<sup>R</sup> either by directly interacting with CCYP<sup>R</sup> or by acting on components of the biological pathway in which CCYP<sup>R</sup> participates.

An "allelic variant" is an alternative form of the gene encoding CCYP<sup>R</sup>. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CCYP<sup>R</sup> include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CCYP<sup>R</sup> or



a polypeptide with at least one functional characteristic of CCYPR. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CCYPR, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence

5 encoding CCYPR. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally

equivalent CCYPR. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the

residues, as long as the biological or immunological activity of CCYPR is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged

amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine.

Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

15 The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic

molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid

20 sequence to the complete native amino acid sequence associated with the recited protein molecule. "Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well

known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CCYPR. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small

25 molecules, or any other compound or composition which modulates the activity of CCYPR either by directly interacting with CCYPR or by acting on components of the biological pathway in which CCYPR participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments

30 thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CCYPR polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or

oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the

translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin,

35 thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that





makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CCYPR, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CCYPR or fragments of CCYPR may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap



(University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

5 "Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue		Conservative Substitution	
10	Ala	Gly, Ser	
	Arg	His, Lys	
	Asn	Asp, Gln, His	
	Asp	Asn, Glu	
	Cys	Ala, Ser	
	Gln	Asn, Glu, His	
	Glu	Asp, Gln, His	
	Gly	Ala	
	His	Asn, Arg, Gln, Glu	
	Ile	Leu, Val	
20	Leu	Ile, Val	
	Lys	Arg, Gln, Glu	
	Met	Leu, Ile	
	Phe	His, Met, Leu, Trp, Tyr	
	Ser	Cys, Thr	
25	Thr	Ser, Val	
	Trp	Phe, Tyr	
	Tyr	His, Phe, Trp	
	Val	Ile, Leu, Thr	

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

35 The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.



A "fragment" is a unique portion of CYP<sub>R</sub> or the polynucleotide encoding CYP<sub>R</sub> which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. For example, to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:55-108 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:55-108, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:55-108 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:55-108 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:55-108 and the region of SEQ ID NO:55-108 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-54 is encoded by a fragment of SEQ ID NO:55-108. A fragment of SEQ ID NO:1-54 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-54. For example, a fragment of SEQ ID NO:1-54 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-54. The precise length of a fragment of SEQ ID NO:1-54 and the region of SEQ ID NO:1-54 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence. "Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.



5	<p>Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.</p> <p>Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <a href="http://www.ncbi.nlm.nih.gov/gorf/b12.html">http://www.ncbi.nlm.nih.gov/gorf/b12.html</a>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:</p> <p><i>Matrix: BLOSUM62</i></p> <p><i>Reward for match: 1</i></p> <p><i>Penalty for mismatch: -2</i></p> <p><i>Open Gap: 5 and Extension Gap: 2 penalties</i></p> <p><i>Gap x drop-off: 50</i></p> <p><i>Expect: 10</i></p> <p><i>Word Size: 11</i></p> <p><i>Filter: on</i></p>
20	<p>Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.</p> <p>Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <a href="http://www.ncbi.nlm.nih.gov/gorf/b12.html">http://www.ncbi.nlm.nih.gov/gorf/b12.html</a>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:</p> <p><i>Matrix: BLOSUM62</i></p> <p><i>Reward for match: 1</i></p> <p><i>Penalty for mismatch: -2</i></p> <p><i>Open Gap: 5 and Extension Gap: 2 penalties</i></p> <p><i>Gap x drop-off: 50</i></p> <p><i>Expect: 10</i></p> <p><i>Word Size: 11</i></p> <p><i>Filter: on</i></p>
35	<p>Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.</p>





Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

5 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

10 Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62  
Open Gap: 11 and Extension Gap: 1 penalties  
Gap x drop-off: 50  
Expect: 10  
Word Size: 3  
Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

35 "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.



The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, *Molecular Cloning: A Laboratory Manual*, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%.

Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions

will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides. The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A

hybridization complex may be formed in solution (e.g.,  $C_{0t}$  or  $R_{0t}$  analysis) or formed between one



nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively. "Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

10 An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CCYPFR which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CCYPFR which is useful in any of the antibody production methods disclosed herein or known in the art.

15 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

20 The term "modulate" refers to a change in the activity of CCYPFR. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CCYPFR.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

25 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript

35 elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an CCYPFR may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in



the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CCYPR.

"Probe" refers to nucleic acid sequences encoding CCYPR, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

"Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al., 1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriiming library," in which

sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific





Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotide and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

10 A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter

sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell. Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

20 A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

30 An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding CCYPR, or fragments thereof, or CCYPR itself, may comprise a bodily fluid; an extract from a cell; chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.



- The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.
- The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.
- A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.
- "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.
- A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.
- "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time. A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be



introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

5 A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

10 A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

30 The invention is based on the discovery of new human cell cycle and proliferation proteins (CCYPR), the polynucleotides encoding CCYPR, and the use of these compositions for the diagnosis, treatment, or prevention of immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding

CCYPR. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the

35 polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte

clones in which nucleic acids encoding each CCYPR were identified, and column 4 shows the cDNA

libraries from which these clones were isolated. Column 5 shows Incyte clones and their



WO 01/07471  
PCT/US00/19948

5 The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis along with relevant citations, all of which are expressly incorporated by reference herein in their entirety; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

10 assemble the consensus nucleotide sequence of each CCYPR and are useful as fragments in hybridization technologies.

15 The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding CCYPR. The first column of Table 3 lists the nucleotide SEQ ID NOS. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:55-108 and to distinguish between SEQ ID NO:55-108 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express CCYPR as a fraction of total tissues expressing CCYPR. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing CCYPR as a fraction of total tissues expressing CCYPR. Column 5 lists the vectors used to subclone each cDNA library. Of particular note is the expression of SEQ ID NO:66 in inflammatory tissues. It should be noted that SEQ ID NO:76 was found to be expressed predominantly in nervous tissue.

20 The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding CCYPR were isolated. Column 1 references the nucleotide SEQ ID NOS, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2. SEQ ID NO:61 maps to chromosome 5 within the interval from 141.40 to 142.60 centiMorgans. This interval also contains gene(s) and/or EST(s) associated with corneal dystrophy and deafness.

25 SEQ ID NO:73 maps to chromosome 2 within the interval from 73.80 to 83.50 centiMorgans. This interval also contains gene(s) and/or EST(s) associated with hereditary nonpolyposis colorectal carcinoma and Muir-Torre syndrome. SEQ ID NO:74 maps to chromosome 19 within the interval from 41.70 to 58.70 centiMorgans. SEQ ID NO:75 maps to chromosome 17 within the interval from

30





62.90 to 64.20 centiMorgans. This interval also contains gene(s) and/or EST(s) located within the human breast cancer (BRCA1) gene region. SEQ ID NO:76 maps to chromosome 1 within the interval from 143.30 to 153.90 centiMorgans, to chromosome 3 within the interval from 156.20 to 160.00 centiMorgans, and to chromosome X within the interval from 112.80 to 139.40 centiMorgans. The interval on chromosome X from 112.80 to 139.40 centiMorgans also contains gene(s) and/or EST(s) associated with X-linked agammaglobulinaemia.

SEQ ID NO:77 maps to chromosome 23 within the interval from 173.60 to 179.80 centiMorgans, and to chromosome 11 within the interval from 136.90 centiMorgans to q-terminus. SEQ ID NO:78 maps to chromosome 3 within the interval from 200.00 to 213.70 centiMorgans. SEQ ID NO:81 maps to chromosome 7 within the interval from 167.60 centiMorgans to q-terminus. SEQ ID NO:90 maps to chromosome 2 within the interval from 236.10 to 240.20 centiMorgans, to chromosome 3 within the interval from 16.50 to 43.00 centiMorgans, and to chromosome 6 within the interval from 22.40 to 40.70 centiMorgans. SEQ ID NO:98 maps to chromosome 8 within the interval from 40.30 to 60.00 centiMorgans. SEQ ID NO:100 maps to chromosome 14 within the interval from 95.50 to 103.70 centiMorgans, and to chromosome 6 within the interval from 158.50 centiMorgans to q-terminus. SEQ ID NO:104 maps to chromosome 18 within the interval from 32.40 to 42.70 centiMorgans. SEQ ID NO:105 maps to chromosome 19 within the interval from 69.90 to 81.20 centiMorgans.

The invention also encompasses CCYPR variants. A preferred CCYPR variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CCYPR amino acid sequence, and which contains at least one functional or structural characteristic of CCYPR.

The invention also encompasses polynucleotides which encode CCYPR. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:55-108, which encodes CCYPR. The polynucleotide sequences of SEQ ID NO:55-108, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding CCYPR. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CCYPR. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:55-108 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of



WO 01/07471 PCT/US00/19948

acid sequence which contains at least one functional or structural characteristic of CCYPR. It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CCYPR, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CCYPR, and all such variations are to be considered as being specifically disclosed.

10 Although nucleotide sequences which encode CCYPR and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CCYPR under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CCYPR or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CCYPR and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

20 The invention also encompasses production of DNA sequences which encode CCYPR and CCYPR derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CCYPR or any fragment thereof.

25 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:55-108 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

30 Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Tag polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably,



sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) *Short Protocols in Molecular Biology*, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) *Molecular Biology and Biotechnology*, Wiley VCH, New York NY, pp. 856-853.)

10 The nucleic acid sequences encoding CCYPB may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)

15 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme

20 digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

30 When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

35 Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary



sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CCYPR may be cloned in recombinant DNA molecules that direct expression of CCYPR, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CCYPR.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CCYPR-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CCYPR, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CCYPR may be synthesized, in whole or in part,





WO 01/07471  
 PCT/US00/19948

using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.)

Alternatively, CCYPR itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of CCYPR, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

15 In order to express a biologically active CCYPR, the nucleotide sequences encoding CCYPR or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CCYPR. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CCYPR. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CCYPR and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used.

30 (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CCYPR and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)



A variety of expression vector/host systems may be utilized to contain and express sequences encoding CCYPR. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; Cornuzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al., (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CCYPR. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CCYPR can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CCYPR into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of CCYPR are needed, e.g. for the production of antibodies, vectors which direct high level expression of CCYPR may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CCYPR. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such



vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, *supra*; and Scorer, *supra*.)

Plant systems may also be used for expression of CCYPR. Transcription of sequences encoding CCYPR may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, *supra*; Broglie, *supra*; and Winter, *supra*.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CCYPR may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CCYPR in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CCYPR in cell lines is preferred. For example, sequences encoding CCYPR can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apv* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic,



or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorosulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g.,

5 Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to 10 quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CCYP<sub>R</sub> is inserted within a marker gene sequence, transformed cells containing sequences encoding CCYP<sub>R</sub> can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CCYP<sub>R</sub> under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CCYP<sub>R</sub> and that express CCYP<sub>R</sub> may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences. Immunological methods for detecting and measuring the expression of CCYP<sub>R</sub> using either 25 specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CCYP<sub>R</sub> is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunocytochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CCYP<sub>R</sub> include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.





Alternatively, the sequences encoding CCYPFR, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CCYPFR may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CCYPFR may be designed to contain signal sequences which direct secretion of CCYPFR through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity.

Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CCYPFR may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CCYPFR protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CCYPFR activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CCYPFR encoding sequence and the heterologous



protein sequence, so that CCYPR may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CCYPR may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

CCYPR of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CCYPR. At least one and up to a plurality of test compounds may be screened for specific binding to CCYPR. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of CCYPR, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, Coligan, J.E. et al. (1991) *Current Protocols in Immunology* 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which CCYPR binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express CCYPR, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing CCYPR or cell membrane fractions which contain CCYPR are then

contacted with a test compound and binding, stimulation, or inhibition of activity of either CCYPR or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CCYPR, either in solution or affixed to a solid support, and detecting the binding of CCYPR to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a

labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

CCYPR of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CCYPR. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CCYPR activity, wherein CCYPR is combined with at least one test compound, and the activity of



CCYP<sub>R</sub> in the presence of a test compound is compared with the activity of CCYP<sub>R</sub> in the absence of the test compound. A change in the activity of CCYP<sub>R</sub> in the presence of the test compound is indicative of a compound that modulates the activity of CCYP<sub>R</sub>. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising CCYP<sub>R</sub> under conditions suitable for CCYP<sub>R</sub> activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CCYP<sub>R</sub> may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CCYP<sub>R</sub> or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell

blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CCYP<sub>R</sub> may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding CCYP<sub>R</sub> can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CCYP<sub>R</sub> is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CCYP<sub>R</sub>, e.g., by secreting CCYP<sub>R</sub> in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-



## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists

between regions of CCYPR and cell cycle and proliferation proteins. In addition, the expression of CCYPR is closely associated with inflammation, trauma, cell proliferation and cancer. Therefore, CCYPR appears to play a role in immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer. In the treatment of disorders associated with increased CCYPR expression or activity of CCYPR. In the treatment of disorders associated with decreased CCYPR expression or activity, it is desirable to increase the expression or activity of CCYPR.

10 Therefore, in one embodiment, CCYPR or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR. Examples of such disorders include, but are not limited to, an immune disorder

such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxicins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, mixed connective tissue disorder (MCTD), multiple sclerosis, myasthenia

20 gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, and hematopoietic cancer including

30 Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, lymphoma, leukemia, and myeloma; a developmental disorder such as renal tubular acidosis, anemia, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidosis, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea

35 and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, sensorineural hearing loss, and disorders of immune cell activation; a cell signaling disorder





- including endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalcemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkaldosis; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, including tubal disease, ovulatory defects, and endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, teratogenesis, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and fibrocystic breast disease; and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5  $\alpha$ -reductase, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.
- 30 In another embodiment, a vector capable of expressing CCYPR or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR including, but not limited to, those described above.
- In a further embodiment, a pharmaceutical composition comprising a substantially purified CCYPR in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR including, but not limited to, those provided above.



In still another embodiment, an agonist which modulates the activity of CCYP<sub>R</sub> may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYP<sub>R</sub> including, but not limited to, those listed above.

In a further embodiment, an antagonist of CCYP<sub>R</sub> may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CCYP<sub>R</sub>. Examples of such disorders include, but are not limited to, those immune, developmental, and cell signaling disorders and cell proliferative disorders including cancer, described above. In one aspect, an antibody which specifically binds CCYP<sub>R</sub> may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CCYP<sub>R</sub>. In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CCYP<sub>R</sub> may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CCYP<sub>R</sub> including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CCYP<sub>R</sub> may be produced using methods which are generally known in the art. In particular, purified CCYP<sub>R</sub> may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CCYP<sub>R</sub>. Antibodies to CCYP<sub>R</sub> may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CCYP<sub>R</sub> or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronics

polys, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (*Bacillus Calmette-Guérin*) and *Corynebacterium parvum* are especially preferable. It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CCYP<sub>R</sub> have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or

fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches



of CCYPR amino acids may be fused with those of another protein, such as K<sub>L</sub>H, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CCYPR may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CCYPR-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for CCYPR may also be generated. For example, such fragments include, but are not limited to, F(ab)<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab)<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CCYPR and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CCYPR epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CCYPR. Affinity is expressed as an association constant, K<sub>a</sub>, which is defined as the molar concentration of CCYPR-antibody complex



WO 01/07471  
PCT/US00/19948

The  $K_d$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CCYPR epitopes, represents the average affinity, or avidity, of the antibodies for CCYPR. The  $K_d$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular CCYPR epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_d$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the CCYPR-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_d$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunoprecipitation and similar procedures which ultimately require dissociation of CCYPR, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CCYPR-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al., *supra*.)

In another embodiment of the invention, the polynucleotides encoding CCYPR, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CCYPR. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CCYPR. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other





WO 01/07471  
PCT/US00/19948

systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding CCYPR may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaise, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and Somia, N. (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Balitmore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in CCYPR expression or regulation causes disease, the expression of CCYPR from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in

CCYPR are treated by constructing mammalian expression vectors encoding CCYPR and introducing these vectors by mechanical means into CCYPR-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vivo* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of CCYPR include, but are not limited to, the pCDNA 3.1, EPITAG, pRCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), pCMV-SCRIPT, pCMV-TAG, pEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CCYPR may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the



tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Biau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Biau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CCYPR from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transfection is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

15 In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CCYPR expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CCYPR under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are

commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant")

discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CCYPR to cells which have one or more genetic abnormalities with respect



to the expression of CCYPR. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Cséte, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CCYPR to target cells which have one or more genetic abnormalities with respect to the expression of CCYPR. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CCYPR to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 69:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CCYPR to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotech. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CCYPR into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CCYPR-coding RNAs and the synthesis of high levels of CCYPR in vector transduced cells. While



alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of CCYPR into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triple DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.L. Carr, *Molecular and Immunologic Approaches*, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CCYPR.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding CCYPR. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into





cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible

modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3'

ends of the molecule, or the use of phosphorothioate or 2'-O-methyl rather than phosphodiesterase

5 linkages within the backbone of the molecule. This concept is inherent in the production of PNAs

and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine,

quosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine,

cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous

endonucleases.

10 An additional embodiment of the invention encompasses a method for screening for a

compound which is effective in altering expression of a polynucleotide encoding CCYPR.

Compounds which may be effective in altering expression of a specific polynucleotide may include,

but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming

oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-

15 macromolecular chemical entities which are capable of interacting with specific polynucleotide

sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or

promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased

CCYPR expression or activity, a compound which specifically inhibits expression of the

polynucleotide encoding CCYPR may be therapeutically useful, and in the treatment of disorders

20 associated with decreased CCYPR expression or activity, a compound which specifically promotes

expression of the polynucleotide encoding CCYPR may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in

altering expression of a specific polynucleotide. A test compound may be obtained by any method

commonly known in the art, including chemical modification of a compound known to be effective in

25 altering polynucleotide expression; selection from an existing, commercially-available or proprietary

library of naturally-occurring or non-natural chemical compounds; rational design of a compound

based on chemical and/or structural properties of the target polynucleotide; and selection from a

library of chemical compounds created combinatorially or randomly. A sample comprising a

polynucleotide encoding CCYPR is exposed to at least one test compound thus obtained. The sample

30 may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted

biochemical system. Alterations in the expression of a polynucleotide encoding CCYPR are assayed

by any method commonly known in the art. Typically, the expression of a specific nucleotide is

detected by hybridization with a probe having a nucleotide sequence complementary to the sequence

of the polynucleotide encoding CCYPR. The amount of hybridization may be quantified, thus

35 forming the basis for a comparison of the expression of the polynucleotide both with and without

exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide



exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific

polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Amdt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys. An additional embodiment of the invention relates to the administration of a pharmaceutical composition which generally comprises an active ingredient formulated with a pharmaceutical acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of *Remington's Pharmaceutical Sciences* (Maack Publishing, Easton PA). Such pharmaceutical compositions may consist of CCYPR, antibodies to CCYPR, and mimetics, agonists, antagonists, or inhibitors of CCYPR.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Pharmaceutical compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of



administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of pharmaceutical compositions may be prepared for direct intracellular delivery of macromolecules comprising CCYPR or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CCYPR or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CCYPR or fragments thereof, antibodies of CCYPR, and agonists, antagonists or inhibitors of CCYPR, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.



Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind CCYPR may be used for the diagnosis of disorders characterized by expression of CCYPR, or in assays to monitor patients being treated with CCYPR or agonists, antagonists, or inhibitors of CCYPR. Antibodies useful for diagnostic assays for CCYPR include methods which utilize the antibody and a label to detect CCYPR in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CCYPR, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CCYPR expression. Normal or standard values for CCYPR expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to CCYPR under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CCYPR expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CCYPR may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CCYPR may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CCYPR, and to monitor regulation of CCYPR levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CCYPR or closely related molecules may be used to identify nucleic acid sequences which encode CCYPR. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CCYPR, allelic variants, or related





sequences.

Probes may also be used for the detection of related sequences, and may have at least 50%

sequence identity to any of the CCYPR encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:55-108 or

5 from genomic sequences including promoters, enhancers, and introns of the CCYPR gene.

Means for producing specific hybridization probes for DNAs encoding CCYPR include the

cloning of polynucleotide sequences encoding CCYPR or CCYPR derivatives into vectors for the

production of mRNA probes. Such vectors are known in the art, are commercially available, and may

be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA

10 polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a

variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels,

such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CCYPR may be used for the diagnosis of disorders

associated with expression of CCYPR. Examples of such disorders include, but are not limited to, an

15 immune disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome

(AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis,

amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia,

autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dysrophy

(APCED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic

20 dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with

lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,

Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal

hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, mixed connective tissue

disorder (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation,

25 myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis,

Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis,

systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic

purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and

30 extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections,

trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a developmental

disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism,

Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms'

tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome,

35 myelodysplastic syndrome, hereditary mucocutaneous dysplasia, hereditary keratodermas, hereditary

neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism,

hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida,



- disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalcemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, teratogenesis, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and fibrocystic breast disease; and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5  $\alpha$ -reductase, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding CCYPR may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CCYPR expression. Such qualitative or quantitative methods are well known in the art.
- In a particular aspect, the nucleotide sequences encoding CCYPR may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide



WO 01/07471  
 PCT/US00/19948

sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CCYPR in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

10 In order to provide a basis for the diagnosis of a disorder associated with expression of CCYPR, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CCYPR, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values

15 obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

20 Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

25 With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

30 Additional diagnostic uses for oligonucleotides designed from the sequences encoding CCYPR may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding CCYPR, or a fragment of a polynucleotide complementary to the polynucleotide encoding CCYPR, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or

35 quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding CCYPR may be used to detect single nucleotide polymorphisms (SNPs). SNPs are



substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded

conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP,

oligonucleotide primers derived from the polynucleotide sequences encoding CCYPR are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis

methods, termed *in silico* SNP (iSSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry

using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA). Methods which may also be used to quantify the expression of CCYPR include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the

polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large

numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and

monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly

effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for CCYPR, or CCYPR or fragments thereof may





be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhammer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological



sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CCYPs to quantify the levels of CCYP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoz, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor



correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding CCYPR may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J.



WO 01/07471  
PCT/US00/19948

et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Urich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CCYP<sub>R</sub> on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

*In situ* hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CCYP<sub>R</sub>, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CCYP<sub>R</sub> and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geyesen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CCYP<sub>R</sub>, or fragments thereof, and washed. Bound CCYP<sub>R</sub> is then detected by methods well known in the art. Purified CCYP<sub>R</sub> can also be coated directly onto plates for use in the aforementioned drug screening techniques.

Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CCYP<sub>R</sub> specifically compete with a test compound for binding





CCYPFR. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CCYPFR.

In additional embodiments, the nucleotide sequences which encode CCYPFR may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder

10 of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/145,075, U.S. Ser. No. 60/153,129, and U.S. Ser. No. 60/164,647, are hereby expressly incorporated by reference.

## 15 EXAMPLES

### I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIzol (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the

recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic

oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column

chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs



were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pCDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microlidispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the



art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire

into full length polynucleotide sequences using programs based on Phred, Phrap, and Conseq, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying

against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:55-108. Fragments from about 20 to about 400 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

#### IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel, 1995, *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESeq (Incyte Genomics). This analysis is



much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\text{BLAST Score} \times \text{Percent Identity} = \frac{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding CCYPR occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

## V. Chromosomal Mapping of CCYPR Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:55-108 were compared with sequences from the Incyte LIFESeq database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:55-108 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Génethon were used to determine if any of the clustered sequences





had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

The genetic map locations of SEQ ID NO:61, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:81, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:104, and SEQ ID NO:105 are described in The

Invention as ranges, or intervals, of human chromosomes. More than one map location is reported for SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:90, and SEQ ID NO:100, indicating that previously mapped sequences having similarity, but not complete identity, to SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:90, and SEQ ID NO:100 were assembled into their respective clusters. The map position

of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM

distances are based on genetic markers mapped by Génethon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

## VI. Extension of CCYPR Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:55-108 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one

extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the



WO 01/07471 PCT/US00/19948

alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

5 The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

10 The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

20 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

30 In like manner, the polynucleotide sequences of SEQ ID NO:55-108 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

## VII. Labeling and Use of Individual Hybridization Probes

35 Hybridization probes derived from SEQ ID NO:55-108 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide



WO 01/07471  
PCT/US00/19948

fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T $_4$  polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10 $^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nyttran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

**VIII. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Sचना (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Sचना, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.



Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 μl volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37 °C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85 °C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate





buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60 °C followed by washes in 0.2% SDS and distilled water as before.

#### Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and

5 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample

mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered

with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just

slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the

addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is

10 incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash

buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X

SSC), and dried.

#### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an

15 Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines

at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is

focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide

containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-

scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a

20 resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.

Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477,

Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate

25 filters positioned between the array and the photomultiplier tubes are used to filter the signals. The

emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is

typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source,

although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a

cDNA control species added to the sample mixture at a known concentration. A specific location on

30 the array contains a complementary DNA sequence, allowing the intensity of the signal at that

location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples

from different sources (e.g., representing test and control cells), each labeled with a different

fluorophore, are hybridized to a single array for the purpose of identifying genes that are

differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the

35 two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital



(A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical cross-talk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum. A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### IX. Complementary Polynucleotides

Sequences complementary to the CCYPR-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CCYPR. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CCYPR. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CCYPR-encoding transcript.

#### X. Expression of CCYPR

Expression and purification of CCYPR is achieved using bacterial or virus-based expression systems. For expression of CCYPR in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*lac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CCYPR upon induction with isopropyl beta-D-

thiogalactopyranoside (IPTG). Expression of CCYPR in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CCYPR by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (SF9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K.



et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther.

7:1937-1945.)

In most expression systems, CCYPR is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-

kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on

immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from

CCYPR at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunofluorescence

10 purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman

Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate

resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995,

*supra*, ch. 10 and 16). Purified CCYPR obtained by these methods can be used directly in the assays

shown in Examples XI and XV.

## 15 XI. Demonstration of CCYPR Activity

An assay for CCYPR activity measures cell proliferation as the amount of newly initiated

DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding CCYPR is

transfected into quiescent 3T3 cultured cells using methods well known in the art. The transiently

transfected cells are then incubated in the presence of [<sup>3</sup>H]thymidine, a radioactive DNA precursor.

20 Where applicable, varying amounts of CCYPR ligand are added to the transfected cells.

Incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA is measured over an appropriate time

interval, and the amount incorporated is directly proportional to the amount of newly synthesized

DNA and CCYPR activity.

## XII. Functional Assays

25 CCYPR function is assessed by expressing the sequences encoding CCYPR at

physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a

mammalian expression vector containing a strong promoter that drives high levels of cDNA

expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1

30 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant

vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic

cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid

containing sequences encoding a marker protein are co-transfected. Expression of a marker protein

provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor

of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green

35 Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM),

an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or

CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects

.

.

.

.

and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CCYPR on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CCYPR and either CD64 or CD64-GFP. 10 CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CCYPR and other genes of interest can be analyzed by northern analysis or microarray techniques.

### XIII. Production of CCYPR Specific Antibodies

CCYPR substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CCYPR amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CCYPR activity by, for example, binding the peptide or CCYPR to a substrate, blocking with 1% BSA, reacting with rabbit antiserum, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

### XIV. Purification of Naturally Occurring CCYPR Using Specific Antibodies

Naturally occurring or recombinant CCYPR is substantially purified by immunoaffinity chromatography using antibodies specific for CCYPR. An immunoaffinity column is constructed by covalently coupling anti-CCYPR antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is





blocked and washed according to the manufacturer's instructions.

Media containing CCYPR are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CCYPR (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CCYPR binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CCYPR is collected.

#### XV. Identification of Molecules Which Interact with CCYPR

CCYPR, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CCYPR, washed, and any wells with labeled CCYPR complex are assayed. Data obtained using different concentrations of CCYPR are used to calculate values for the number, affinity, and association of CCYPR with the candidate molecules.

Alternatively, molecules interacting with CCYPR are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech). CCYPR may also be used in the PATHCALLING process (Curagen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.



Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	55	116462	KIDNOT01	116462H1 (KIDNOT01), 116462R1 (KIDNOT01), 116462X304D1 (KIDNOT01), 1500439F6 (SINTBST01), 2369977F6 (ADRENOT07)
2	56	1210462	BRSTNOT02	260707H1 (HNT2RAT01), 1210462H1 (BRSTNOT02), 1458882F6 (COLNFET02), 1841248T6 (COLNNOT07), 2378362H1 (ISLTNOT01), 3728643F6 (SMCCNON03)
3	57	1305252	PLACNOT02	794067R6 (OVARNOT03), 871989R1 (LUNGAST01), 1235253F1 (LUNGFET03), 1305252F6 (PLACNOT02), 1305252H1 (PLACNOT02), 1703258T6.comp (DUODNOT02), 2678307H1.comp (OVARUT07), 3221088H1.comp (COLNNON03), 3647280H1 (ENDINOT01)
4	58	1416289	BRAINOT12	639958R6 (BRSTNOT03), 861752H1 (BRAITUT03), 1416289H1 (BRAINOT12), 1416289X310B1 (BRAINOT12), 1416289X310D2 (BRAINOT12), 1947451R6 (PITUNOT01)
5	59	1558289	SPLNNOT04	1558289H1 (SPLNNOT04), 1852450T6 (LUNGFET03), 2396092F6 (THPIAZT01), 2593267F6 (LUNGNOT22), 2632784F6 (COLANTUT15)
6	60	1577739	LNODNOT03	181266R1 (PLACNOB01), 1577739H1 (LNODNOT03), 4180022T6 (SINITUT03), 4597046H1 (COLSTUT01), 4860616H1 (PROSTUT09), 4991290H1 (LIVRTUT11), 5059810H1 (CONDTUT02)
7	61	1752768	LIVRTUT01	256106R1 (HMT2RAT01), 258814H1 (HMT2RAT01), 1312247F1 (COLNFET02), 1344279T6 (PROSNOT11), 1350089H1 (LATRTUT02), 1440718F6 (THYRNOT03), 1752768F6 (LIVRTUT01), 1752768H1 (LIVRTUT01), 1752768T6 (LIVRTUT01), 2079106F6 (ISLTNOT01), SBYA00612U1
8	62	1887228	BLADTUT07	080294F1 (SYNORAB01), 140055F1 (TLYMNOR01), 285207X42 (EOSIHET02), 516882R6 (MMLRIDT01), 1217892T1 (NEUTGMT01), 1887228H1 (BLADTUT07), 4323029H1 (TLYMUNT01)
9	63	1988468	LUNGAST01	072147R6 (THPIPEB01), 496297H1 (HMT2NOT01), 1362109F6 (LUNGNOT12), 1726095F6 (PROSNOT14), 1726095T6 (PROSNOT14), 1988468H1 (LUNGAST01), 1988468T6 (LUNGAST01), 2232471F6 (PROSNOT16)
10	64	2049176	LIVRFET02	2049176H1 (LIVRFET02), 2049176T6 (LIVRFET02), 2049176X321D1 (LIVRFET02)
11	65	2686765	LUNGNOT23	1502858F6 (BRAITUT07), 1956694X315D1 (CONNNOT01), 2022628X307D1 (CONNNOT01), 2686765F6 (LUNGNOT23), 2686765H1 (LUNGNOT23), 2864555H1 (KIDNNOT20), 2887609F6 (SINJNOT02), 3381980H1 (ESOGNOT04)
12	66	3215187	TESTNOT07	151135R6 (FIBRAGT01), 3215187F6 (TESTNOT07), 3215187H1 (TESTNOT07)
13	67	3500375	PROSTUT13	860585R1 (BRAITUT03), 1318501F1 (BLADNOT04), 1419126F1 (KIDNNOT09), 1483246F6 (CORPNOT02), 2238114T6 (PANCTUT02), 2272329H1 (PROSNON01), 3209746F7 (BLADNOT08), 3403213H1 (ESOGNOT03), 4176619H1 (BRAINOT22), 4614606H1 (BRAYDIT01)

.

.

.

.

Table 1 (cont.)

Polypeptide Seq ID NO:	Nucleotide Seq ID NO:	Clone ID	Library	Fragments
14	68	5080410	LNODNOT11	1270372X300D1 (BRAINOT09), 3460603H1 (293TF1T01), 5080410H1 (LNODNOT11)
15	69	5218248	BRSTNOT35	1808748X15C1 (PROSTUT12), 1808748X16C1 (PROSTUT12), 3391884H1 (LUNGNOT28)
16	70	058336	MUSCNOT01	058336H1 (MUSCNOT01), 058336T6 (MUSCNOT01), 92206766, 92069225
17	71	1511488	LUNGNOT14	1436265F1 (PANCNOT08), 1511488H1 (LUNGNOT14), 1511488T6 (LUNGNOT14), 1850020F6 (LUNGEFET03)
18	72	1638819	UTRSNOT06	1282638T1 (COLANOT16), 1638819F6 (UTRSNOT06), 1638819H1 (UTRSNOT06), 3597071H1 (FIBPNOT01), SBRA03813D1, SBRA04133D1, SBRA03785D1
19	73	1655123	PROSTUT08	1271351F1 (TESTTUT02), 1353234F1 (LATRTUT02), 1655123H1 (PROSTUT08), 2132186R6 (OVARNOT03), 3296525H1 (TLXJINT01), 3354010H1 (PROSNOT28), 3741838F6 (MENTNOT01), 3741838T6 (MENTNOT01), SXAF03528V1
20	74	2553926	THYMNOT03	403261F1 (TMLR3D1T01), 1869739F6 (SKINBIT01), 2197242T6 (SPINFEET02), 2553926H1 (THYMNOT03), 2553956T6 (THYMNOT03), 3935528H1 (PROSTUT09), 5263918F6 (CONDTUT02)
21	75	2800717	PENCNOT01	411179F1 (BRSTNOT01), 415284R1 (BRSTNOT01), 1458971F1 (COLNEET02), 1600810H1 (BLADNOT03), 1622005F6 (BRAITUT13), 2173076F6 (ENDCNOT03), 2520087F6 (BRAITUT21), 2800717H1 (PENCNOT01), 5184583H1 (LUNGTMOT03), 5435834H1 (SPINNOT17), 5872662H1 (COLTDIT04)
22	76	5664154	BRAUNOT01	181534F1 (PLACNOB01), SCHAO0262V1
23	77	017900	HUVELPB01	017900H1 (HUVELPB01), 092858F1 (HYPONOB01), 1353543F1 (LATRTUT02), 1353543F6 (LATRTUT02), 1428464F1 (SINIBST01), 91616429
24	78	035102	HUVENOB01	035102H1 (HUVENOB01), 077722R1 (SYNORAB01), 995133H1 (KIDNTUT01), 1356968T6 (LUNGNOT09), 1963135R6 (BRSTNOT04), 2659921F6 (LUNGUTUT09), 3110603H1 (BRSTNOT17)
25	79	259983	HNT2RAT01	259131R1 (HNT2RAT01), 259983H1 (HNT2RAT01), 268205R1 (HNT2NOT01), 1305726F1 (PLACNOT02)
26	80	926810	BRAINOT04	926810H1 (BRAINOT04), 3490378T6 (EPIGNOT01), 4774848H1 (BRAONOT01), SBIA01080D1, SBIA04006D1, SBIA02273D1, SBIA01121D1
27	81	1398816	BRAITUT08	056398F1 (FIBRNOT01), 1252138F2 (LUNGEFET03), 1294556T1 (PGANNOT03), 1398816H1 (BRAITUT08), 1545328R1 (PROSTUT04)



Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
28	82	1496820	PROSNON01	996673H1 (KIDNTUT01), 1496820H1 (PROSNON01), 2368484F6 (ADRENOT07), 3071781X303D1 (UTRSNOR01), 3071781X307B1 (UTRSNOR01), 3071781X316B2 (UTRSNOR01), 3071781X316D3 (UTRSNOR01)
29	83	1514559	PANCTTUT01	155768H1 (THPIPLB02), 1229952H1 (BRAITUT01), 1337018X11 (COLNNOT13), 1360361H1 (LUNGNOT12), 1365811H1 (SCORNON02), 1514559F6 (PANCTTUT01), 1514559H1 (PANCTTUT01)
30	84	1620092	BRAITUT13	1620092F6 (BRAITUT13), 1620092H1 (BRAITUT13), 1832842H1 (BRAINON01), 1843815R6 (COLNNOT08), 1843815T6 (COLNNOT08)
31	85	1678765	STOMFET01	1678765F6 (STOMFET01), 1678765H1 (STOMFET01), 2640786H1 (LUNGUT08), 3542276F6 (TONSNOT03), 4180591H1 (SINTTUT03), 4183383H1 (LIVRDIR01), 4349212H1 (TLYMTXT01), 4718559H1 (BRAIHCT02), 5023762H1 (OVARNON03), 5332272H1 (KIDNNOT34), 91665766
32	86	1708229	PROSNOT16	388493R1 (THYMNOT02), 1503519F1 (BRAITUT07), 1708229H1 (PROSNOT16), 1725267F6 (PROSNOT14), 3089258F6 (HEAONOT03)
33	87	1806454	SINTNOT13	406723H1 (EOSIHET02), 821556R1 (KERANOT02), 1649621F6 (PROSTTUT09), 1710552H1 (PROSNOT16), 1806454F6 (SINTNOT13), 1806454H1 (SINTNOT13), 2526283H1 (BRAITUT21), 3869969H1 (BMARNOT03)
34	88	1806850	SINTNOT13	270548H1 (HNT2NOT01), 443885R1 (MPHNOT03), 1257235F1 (MENITUT03), 1337438H1 (COLNNOT13), 1351820F1 (LATRTTUT02), 1544066R1 (PROSTTUT04), 1806850F6 (SINTNOT13), 1806850H1 (SINTNOT13), 1984108T6 (LUNGAST01), 2921419H1 (SININOT04), 3109392H1 (BRSTTUT15)
35	89	1851534	LUNGFET03	1851534H1 (LUNGFET03), 2407346R6 (BSTNON02), 2757389R6 (THPIAZS08), 5513454H1 (BRADDIR01), 5629312H1 (PLACFER01)
36	90	1868749	SKINBIT01	1322048F1 (BLADNOT04), 1398330F1 (BRAITUT08), 1437866F6 (PANCNOT08), 1868749F6 (SKINBIT01), 1868749H1 (SKINBIT01), 2279968R6 (PROSNON01), 2684670H1 (LUNGNOT23), 4632232H1 (GBLADIT02), 4951533H2 (ENDVUT01), 5077673H1 (LNODNOT11), 5388496H1 (BRAINOT19)
37	91	1980010	LUNGUT03	127747R1 (TESTNOT01), 357561F1 (PROSNOT01), 357561R1 (PROSNOT01), 918017R1 (BRSTNOT04), 1428117F6 (SINTBST01), 1625080F6 (COLNPOT01), 1720753H1 (BLADNOT06), 1932038F6 (COLNNOT16), 1980010H1 (LUNGUT03), 3112417F6 (BRSTNOT17), 4174704H1 (SINTNOT21), 4238802H1 (SYNWDIT01), 5499543H1 (BRABDIR01), 94337459





Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
38	92	2259032	OVARTUT01	475134H1 (MMLR2DT01), 784284R1 (PROSNOT05), 1264124H1 (SYNORAT05), 1418710F1 (KIDNOT09), 1697570T6 (BLADTUT05), 1874051F6 (LEURNOT02), 2187960T6 (PROSNOT26), 2259032H1 (OVARTUT01), 2259032R6 (OVARTUT01), 3406237H1 (ESOGNOT03), 3441729H1 (PENCNOT06), 3555764H1 (LUNGNOT31), 3728010H1 (SMCCNON03), 3813639H1 (TONSNOT03), 4031501H1 (BRAINTOT23), 4274704H1 (PROSTUT01), 4602450H1 (BRSTNOT07), 93327183
39	93	2359526	LUNGFET05	1667182F6 (BMARNOT03), 2359526H1 (LUNGFET05), 2359526X31DI (LUNGFET05), 2555305F7 (THYMNOT03), 2654667T6 (THYMNOT04), SCHA00290V1, SCHA00266V1, g1748241
40	94	2456494	ENDANOT01	1860223F6 (PROSNOT18), 2456494H1 (ENDANOT01), 2564671H1 (ADRETUT01), 3618339H1 (EPIPNOT01)
41	95	2668536	ESOGTUT02	1513847H1 (PANCTUT01), 1668943F6 (BMARNOT03), 1668943T6 (BMARNOT03), 1721443F6 (BLADNOT06), 2668536H1 (ESOGTUT02), 3438287H1 (PENCNOT05), SBFA00330F1, SCBA05255V1, SCBA01530V1
42	96	2683225	SINIUCT01	196443R6 (KIDNOT02), 1243440R6 (LUNGNOT03), 1604540F6 (LUNGNOT15), 2072837H1 (ISLTNOT01), 2683225F6 (SINIUCT01), 2683225H1 (SINIUCT01), 3647874H1 (ENDINOT01), 4029178H1 (BRAINTOT23)
43	97	2797839	NPOLNOT01	460779T6 (KERANOT01), 782663H1 (MYONOT01), 896898R1 (BRSTNOT05), 1218533H1 (NEUTGMT01), 1312923F6 (BLADTUT02), 2473746F6 (THPINOT03), 2481564H1 (SMCANOT01), 2797839H1 (NPOLNOT01), 3350118H1 (BRAITUT24), 4184264H1 (BRABDIR01), 4401265H1 (TESTTUT03), 4727770H1 (GBLADIT01), 5080203H1 (LNODNOT11), 5524886H1 (LIVRDIR01)
44	98	2959521	ADRENOT09	046696H1 (CORNNOT01), 087727R6 (LIVRNOT01), 138475H1 (LIVRNOT01), 167505H1 (LIVRNOT01), 647975H1 (CARCTXT02), 781084T1 (MYONNOT01), 972191R6 (MUSCNOT02), 1309196H1 (COLNFET02), 2641117H1 (LUNGUT08), 2913953H1 (KIDNTUT15), 2959521H1 (ADRENOT09), 2984654H1 (CARGDIT01), 2985141H1 (CARGDIT01), 3138371H1 (SMCCNOT02), 3386016H1 (ESOGNOT04), 3496187H1 (ADRETUT07), 3614426H1 (EPIPNOT01), 4287819H1 (LIVRDIR01), 5395566H1 (LIVRTUT13), g505101
45	99	3082014	BRAIUNT01	182588H1 (PLACNOB01), 645276R6 (BRSTTUT02), 1497811F1 (SINTBST01), 2051505F6 (LIVRFET02), 3082014H1 (BRAIUNT01), 3464112F6 (293TE2T01), 4603079H1 (BRSTNOT07)
46	100	3520701	LUNGNON03	971201H1 (MUSCNOT02), 1544657R6 (PROSTUT04), 1545570H1 (PROSTUT04), 1671030F6 (BMARNOT03), 1671030T6 (BMARNOT03), 2605263F6 (LUNGUTUT07), 3520701H1 (LUNGNON03), 3520701R6 (LUNGNON03)



Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
47	101	4184320	BRADDIT02	2156956F6 (BRAINOT09), 4184253F6 (BRABDIR01), 4184253T6 (BRABDIR01), 4184320H1 (BRADDIT02), 4252542F6 (BRADDIR01)
48	102	4764233	PLACNOT05	4764233H1 (PLACNOT05), 5634642H1 (PLACFER01), 91148809
49	103	4817352	HELATXT03	4269396 (BLADNOT01), 426993T6 (BLADNOT01), 488301R6 (HNT2AGT01), 3779640H1 (BRSTNOT27), 4817352H1 (HELATXT03)
50	104	5040573	COLHTUT01	1724126F6 (PROSNOT14), 1859337F6 (PROSNOT18), 2026289R6 (KERANOT02), 2026289T6 (KERANOT02), 2122846T6 (BRSTNOT07), 3225302H1 (ADRETUT07), 3322214H1 (PTHYNOT03), 4587178H1 (BRAONOT01), 4601227H1 (BRSTNOT07), 4885408H1 (LUNLTM01), 5040573H1 (COLHTUT01)
51	105	5627029	PLACFER01	967988R1 (BRSTNOT05), 1534642T6 (SPLNNOT04), 1700904F6 (BLADFTUT05), 1846971R6 (COLNNOT09), 2112727R6 (BRAITUT03), 2112727T6 (BRAITUT03), 2205225F6 (SPLNFET02), 2828475H1 (TLYNNOT03), 3439165F6 (PENCNOT06), 3604622H1 (LUNGNOT30)
52	106	5678487	293TF2T01	1258787F6 (MENITUT03), 1522008F1 (BLADFTUT04), 1597992F6 (BLADNOT03), 2057679H1 (BEPINOT01), 2411504H1 (BSTNNON02), 2467956H1 (THYRNOT08), 2739089F6 (OVARNOT09), 2739089T6 (OVARNOT09), 2740762H1 (BRSTTUT14), 2754616H1 (THPLAZS08), 3254971R6 (OVARNTUN01), 3487616H1 (EPIGNOT01), 5678487H1 (293TF2T01)
53	107	5682976	BRAENOT02	350492H1 (LVENNOT01), 825361R1 (PROSNOT06), 879866R1 (THYRNOT02), 1667502F6 (BMARNOT03), 1733323F6 (BRSTTUT08), 1876248T6 (LEUKNOT02), 1963215T6 (BRSTNOT04), 2539188H1 (BONRTUT01), 2896448H1 (KIDNTUT14), 3141553H1 (SMCCNOT02), 3374826F6 (CONNTUT05), 3773427H1 (BRSTNOT25), 3779981H1 (BRSTNOT27), 5682976H1 (BRAENOT02), 5546853H1 (TESTNOC01)
54	108	5992432	FTUBTUT02	645878R6 (BRSTTUT02), 1287660F1 (BRAINOT11), 1287660T6 (BRAINOT11), 1417373F6 (BRAINOT12), 1618868F6 (BRAITUT12), 2269980R6 (UTRSNOT02), 2793117F6 (COLNTUT16), 3246793F6 (BRAINOT19), 3592787H1 (293TF5T01), 5992432H1 (FTUBTUT02), 9821012



Table 2

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
1	145	T10 S93	N15 N38	Signal peptide: M1-Q33 Protein SH3 domain repeat: L8-R99 GLGF signal transduction-related domain: M1-R99		MOTIFS SPSCAN BLAST_PRODOM BLAST_DOMO
2	340	T39 S190 S268 T307 S88 S102 S165 S226 S230 S234 T337		P120 nuclear proliferating cell antigen: N117-K333 Proliferative cell nucleolar protein P120: E26-G293	Proliferating cell nucleolar antigen P120 (g2649749) <u>A. fulgidus</u>	MOTIFS BLAST_PRODOM BLAST_DOMO BLAST_GenBank
3	418	S246 S415 T142 T156 S292 S349 S369 S64 S247 S298	N190 N191 N203 N288 N306		Candidate tumor suppressor p33ING1 (g2829208) <u>H. sapiens</u>	MOTIFS BLAST_GenBank
4	297	T217 T82 S76 S127 S176 T207 S246 Y189	N74	Germ cell-less protein: E96-N297	Germ cell-less protein (g5814404) <u>Mus musculus</u>	MOTIFS BLIMPS_PFAM BLAST_GenBank
5	184	T34 S103 S5 T136	N76		Differentiation factor MDC-3.13 (g3860093) <u>H. sapiens</u>	MOTIFS BLAST_GenBank
6	173	S109 S24 S59 S66 S141 S142 T152			Posterior end mark-5 (g4107015) <u>C. savignyi</u>	MOTIFS BLAST_GenBank



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
7	591	S582 T71 T208 S217 S339 T475 S493 T536 S45 S105 S153 T208 S305 S336 T578 Y93	N374 N425 N534 N585	Signal peptide M1- L64 TPR domain mitosis control E239-P356	Cell division cycle protein 23 homolog (g5541721) <u>A. thaliana</u>	MOTIFS SPSCAN HMMR_PFAM BLAST_DOMO BLAST_GenBank
8	463	T237 S34 T67 T117 T125 S138 T288 T321 S328 S418 T80 S186 S190 S209 S210 T232 T288 S418 T441 S445 Y416	N208	Formin limb deformity: M1-E335	Lymphocyte specific formin related protein (g4101720) <u>M. musculus</u>	MOTIFS BLAST_PRODOM BLAST_DOMO BLAST_GenBank
9	270		N64 N94 N147		Early embryogenesis MRG1 protein (g2570051) <u>M. musculus</u>	MOTIFS BLAST_GenBank
10	255	S180 T49 T53 S97 S152 T201 S210 S23 S97 T145 T216 S225 S228 T231 S242 Y106 Y240		Polyposis locus TB2 homolog: G15-T117 Polyposis locus protein: V13-T117	Similar to polyposis locus protein 1 (g849238) <u>H. sapiens</u>	MOTIFS BLAST_PRODOM BLAST_DOMO BLAST_GenBank
11	533	S227 S412 S505 S7 S17 S65 T349 S442 T29 S72 S89 S358 S442 T446 S505 Y244		TRE oncogene: R56- I277	TRE oncogene- related protein (g2286196) <u>D. melanogaster</u>	MOTIFS BLOCKS_DOMO BLAST_GenBank





Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
12	160	S40		Signal peptide: M1-A30 Transmembrane domain: A6-I29 Cornichon developmental protein: M1-S160	Cornichon-like protein (g4521254) <i>M. musculus</i>	MOTIFS SPSCAN HMMR BLAST_PRODOM BLAST_DOMO BLAST_GenBank
13	531	S195 T196 S357 T45 S172 T199 S212 S322 S465 T495 T45 T241 S255 T279 T319 S328	N244 N401		Cdc 73p (g632679) <i>S. cerevisiae</i>	MOTIFS BLAST_GenBank
14	165	S3 T67 S104			Wolf-Hirschhorn syndrome candidate 2 protein (g3860187) <i>H. sapiens</i>	MOTIFS BLAST_GenBank
15	199	S2 S21 S69 T102 S189			Developmental protein DG118 (g3789911) <i>D. discoideum</i>	MOTIFS BLAST_GenBank
16	168	S141 S55 S61 T79	N77	Signal peptide M1-S61 H-Rev protein homolog P15-K166	g3777529 retinoic acid receptor responder 3 <i>Homo sapiens</i>	BLAST-GenBank SPSCAN BLAST-PRODOM MOTIFS
17	162	S70 S85 T16 T28 T65 T80 T100 S127 Y111			g207250 growth and transformation dependent protein <i>Rattus norvegicus</i>	BLAST-GenBank



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
18	246	T209 S227 T243 T28 S223 S51 S136 S201	N26 N158	Protein cell intergenic region FTSJ K25-K241	<u>g2622903</u> cell division protein J <u>Methanobacterium thermoauto-trophicum</u> <u>g1322234</u> OS-9 precursor <u>Homo sapiens</u>	BLAST-GenBank BLAST-PRODOM BLAST-DMO MOTIFS
19	483	T394 T85 S86 S219 S225 T230 S298 T299 T472 S114 S200 T273 S371 T407 T424 T431		Signal peptide M1-G29 OS-9 precursor L54-E281	<u>g1322234</u> OS-9 precursor <u>Homo sapiens</u>	BLAST-GenBank SPSCAN BLAST-PRODOM MOTIFS
20	280	T129 T6 T102 T119 T181 S250 S46 T72 T84 S262		Signal peptide M1-L28	<u>g3901272</u> ZW10 interactor Zwint <u>Homo sapiens</u>	BLAST-GenBank SPSCAN MOTIFS
21	425	S122 S235 T60 S192 S203 S204 S218 S226 S307 T313 S332 S366 S370 T375 T402 S409 S89 S118 S241 S284 T360 Y399	N190 N311		<u>g455719</u> Activated c-raf oncogenic fusion protein homolog <u>Homo sapiens</u>	BLAST-GenBank
22	128	S3 S107	N42	Prenyl group binding site (CAAX box) C125-P128 Ovarian granulosa cell 13.0 KD protein HGR74 N16-P128	<u>g4580592</u> brain expressed X-linked protein 2 <u>Mus musculus</u>	BLAST-GenBank MOTIFS BLAST-PRODOM
23	113	S88 T20 T37		Biotin-requiring enzyme attachment site: L40-L90	LDOC-1 protein g3869127 ( <u>Homo sapiens</u> ) Nagasaki, K. et al. (1999) Cancer Lett. 140:227-234.	BLAST-GenBank PROFILERSCAN MOTIFS



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
24	308	S95 T79 T98 S184 S246 S251 T55 S184 S226 S294 S300 Y127	N77	Melanoma antigen gene (MAGE) family: M1-Q200, H205- D283, D91-A287	Breast cancer associated gene 1 g4928044 (Homo sapiens) Lurquin, C. et al. (1997) Genomics 46:397-408.	BLAST-GenBank BLAST-PRODOM HMMER-PFAM BLAST-DOMO MOTIFS
25	221	S145 S160 S217 S25 S31 S70 S85 T89 S153 S197 Y34	N139	Annexin VI signature: L86-V95 Sushi domain: T165-C174	Teratocarcinoma expressed gene Tera g1575505 (Mus musculus)	BLAST-GenBank BLIMPS-PRINTS BLIMPS-PFAM MOTIFS
26	402	T344 S39 S78 S109 S237 T269 S273 T376 T381 T383 S11 S49 T89 T344 S364	N76 N107 N171 N362		Paraneoplastic cancer-testis- brain antigen g6179740 (Homo sapiens)	BLAST-GenBank MOTIFS
27	93	S11			Hypoxia inducible gene-1 g4929330 (Homo sapiens)	BLAST-GenBank MOTIFS
28	353	S125 T42 S43 S85 S212 S283 S314 T42 S49 S105 S120 S133 S162 S163 S212 S290	N145 N157 N191	af-4 (FEL protein): S195-K353 E4-Q185	AF5q31 protein g6601438 (Homo sapiens)	BLAST-GenBank BLAST_PRODOM BLAST-DOMO MOTIFS
29	120	T57		Cyclin-dependent kinase inhibitor: D7-P106, M1-N114	Cyclin dependent kinase inhibitor CIP1 g2276312 (Homo sapiens)	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
30	144	S15 S64		Transmembrane domain: I93-I110	Transformation dependent protein g207250 (Rattus norvegicus) N.Glaichenhaus and F.Cuzin (1987) Cell 50:1081-1089.	BLAST-GenBank MOTIFS HMER
31	933	S603 T51 S109 T129 S162 S203 S223 S224 S240 S261 S266 S280 S282 S313 T328 S346 S353 S378 S394 S460 S491 S499 T531 S627 S641 S642 S725 T732 S759 S188 S309 S423 S592 S671 S675 T706 S771 Y856	N107 N238 N639 N883		Replication protein Smp2 g218488 (Saccharomyces cerevisiae) Irie, K. et al. (1993) Mol. Gen. Genet. 6:283-288.	BLAST-GenBank MOTIFS
32	268	S7 T104 T154 S169	N90	Serine-threonine kinase Binder MPS1: L74-I230	Putative mitotic protein (Schizosaccharomyces pombe) g3947877 F.C.Luca and M.Winey (1998) Mol Biol Cell 9:29-46.	BLAST-GenBank BLAST-PRODOM MOTIFS





Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
33	337	T29 S236 T44 T238		Leucine zipper: L259-L280, L266-L287	DNA binding protein g184390 (Homo sapiens) Weitzel, J.N. et al. (1992) Genomics 14:309-319.	BLAST-GenBank MOTIFS
34	565	T17 S34 S61 S66 T138 T142 S174 T238 S245 S265 S436 S466 S527 S106 S205 S218 S258 T297 S314 T325 S463 T470 Y460	N347 N386 N506	F-Box domain: H75-Y123, L82-N95 Disease resistance protein: G254-I270	F-box protein FLR1 g7672734 (Homo sapiens)	BLAST-GenBank HMMER_Pfam BLIMPS-PRINTS MOTIFS
35	228	S200 T47 T62 S78 S107 S188 S192 S206 S200 S205 S213	N36 N94 N225		Predicted WHSC1 protein (Wolf-Hirschhorn syndrome critical region 1) g4378022 (Homo sapiens) Stecc I. et al. (1998) Hum. Mol. Genet. 7:1071-1082.	BLAST-GenBank MOTIFS
36	495	S451 S152 S365 S478 S108 S171 S181 T192 T347 T409 S435 Y86 Y111 Y203			Malignant brain tumor protein 1 (3)mbt g3811111 (Homo sapiens) Koga, H. et al. (1999) Oncogene 18:3799-3809.	BLAST-GenBank MOTIFS



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
37	1336	T635 T769 S902 S10 S32 S33 T76 S95 S156 T298 S313 T427 S467 T579 T626 T642 S661 T668 S680 T699 T729 S774 S834 T859 T915 S944 S959 S961 S997 S1049 T1085 S1132 S1227 T1245 S1249 T48 S94 T169 S224 T352 T379 T389 T475 T696 S867 T883 T889 S940 S961 S1220 Y631	N148 N152 N345 N385 N1213 N1247	Ribosomal protein S14 signature: R1172-N1194 Leucine zipper: L211-L232	Neuroblastoma related protein g4337460 (Homo sapiens)	BLAST-GenBank BLIMPS-PRINTS MOTIFS
38	934	T532 S11 T23 T80 S171 S202 T214 T240 S244 T275 S412 S416 S437 S518 T523 S719 S746 S753 S796 S807 S93 T279 T527 S598 T780	N8 N210 N426	SAP: I92-Q364	Sap2 family putative cell cycle dependent phosphatase g3426127 (Schizosaccharomyces pombe) Luke, M.M. et al. (1996) Mol. Cell Biol. 16:2744-2755.	BLAST-GenBank BLAST-DMO MOTIFS



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
39	515	T72 S122 S175 S272 S277 S305 T420 S422 T432 T79 S139 T189 S215 T316 S457 T486 Y13 Y383	N16 N31 N115	Metastasis-Associated Protein: E65-R230 Leucine zipper: L234-L255	Metastasis associated gene g1008544 (Homo sapiens) Toh, Y. et al. (1995) Gene 159:97-104 Toh, Y, et al. (1994) J Biol. Chem. 269:22958-22963.	BLAST-GenBank BLAST-PRODOM BLIMPS-PRINTS MOTIFS
40	146	S61		Leucine zipper: L5-L26, L12-L33, L19-L40	LD0C1 g3869127 (Homo sapiens)	BLAST-GenBank BLIMPS-PFAM MOTIFS
41	580	S324 S36 S340 S550 S86 T109 T119 T150 T226 S329 S340	N190	Cyclin: H19-K262	Cyclin K g3746549 (Homo sapiens) Edwards, M.C. et al. (1998) Mol. Cell Biol. 18:4291-4300.	BLAST-GenBank BLAST-PRODOM MOTIFS
42	131	S78 T121 T26		Presenilin: Q64-K75	Cell growth regulator DRRI g4322559 (Homo sapiens) G. Thomas and M.N. Hall (1997) Curr. Opin. Cell Biol. 9:782-787.	BLAST-GenBank BLIMPS-PRINTS MOTIFS



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
43	812	S44 S588 S646 S801 S111 S120 S134 T140 S148 S150 S181 T185 S262 S279 S440 T477 S497 T520 T542 T605 S675 S40 T64 T311 T316 T319 T505 S562 S565 T566 T695 S702 S707 S708 T739 T776 S790 Y277	N503 N618	NOL1/NOP2/fmu(sun) family signature: F454-G467, F300-K585, I388-M402, G410-G433, F454-G467, K507-L532, E189-M576 Proliferating Cell Nucleolar Antigen P120: M1-S134, E135- T311, F587-G805	Proliferating cell nuclear protein P120 g287723 (Homo sapiens)	BLAST-GenBank BLAST-PRODOM BLAST-DMO BLIMPS-BLOCKS MOTIFS HMMER-PFAM
44	537	S505 T69 S138 S194 S310 S337 S356 T386 S485 S37 T45 T282	N122 N132 N147	Transmembrane domains: I506-G532, V271-L290, W472-F490	Estrogen induced protein in breast cancer LIV-1 g1256001 (Homo sapiens)	BLAST-GenBank HMMER MOTIFS
45	584	S185 T324 S343 T537 S575 S17 T102 S128 T229 T374 S412 T450	N28	Cytochrome C motif: C283-T288 Metastasis- associated protein MTA1: R19-R143, D144-K321, G340-G483, P432-K555 Leucine zipper: L147-L168	Metastasis associated gene g1008544 (Homo sapiens) Toh, Y. et al. (1995) Gene 159:97-104 Toh, Y. et al. (1994) J. Biol. Chem. 269:22958-22963.	BLAST-GenBank BLAST-PRODOM MOTIFS





Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
46	425	S190 T301 S12 S19 S41 S205 T206 T235 S263 S265 T315 S43 S52 S85 T93 T351 S411 Y422	N275	ML02 mitosis-associated protein: L24-R188, P226-Y245, N308-E408		BLAST-PRODOM MOTIFS
47	255	T9 T147 S237	N144	Melastatin: M1-R172, G199-G255	Melastatin g3047242 (Mus musculus) Duncan, L.M. et al. (1998) Cancer Res. 58:1515-1520.	BLAST-GenBank BLAST-PRODOM MOTIFS
48	111	T30 S2 T8			Melanoma associated antigen GAGE-8 g3511023 (Homo sapiens) Van den Eynde, B. et al. (1995) J. Exp. Med. 182:689-698.	BLAST-GenBank MOTIFS
49	422	T110 T159 S136 S150 T163 T190 S383 T413 S9 T27 S46 S96 T347 S359 S363 S368 Y350		XPMC2 (mitosis associated inducing protein): A236-E402	Mitotic regulator XPMC2 (Xenopus gene which prevents mitotic catastrophe) g595380 (Xenopus laevis) J.Y.Su and J.L.Maller (1995) Mol. Gen. Genet. 246:387-396.	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
50	397	S20 S21 T395 T57 S59 T64 S127 S208 T210 S262 S307 T341 T64 T168 S180 S185 S218 S231 S288 S326	N222 N260	Transmembrane motifs: I361-L380, L24-L44 Cell division control protein: K17-L347	Cell cycle protein CDC1 g550426 (Saccharomyces cerevisidae)	BLAST-GenBank HMMER BLAST-PRODOM MOTIFS
51	800	S56 S448 T721 S760 S48 S84 S111 S119 T146 T189 T235 S250 S265 T275 S321 S335 T392 S448 T466 S474 T562 S596 S598 T626 S686 S3 S4 S65 S89 S107 T123 S348 T398 T402 T716 S730 S738 T743 S789 Y102 Y316 Y569 Y685	N554 N665	Signal peptide: M1-A25 Leucine zipper: L365-L386	SART-1 g4126469 (Mus musculus)	BLAST-GenBank SPSCAN MOTIFS
52	713	S100 T631 S8 T9 S20 T42 T114 T121 T172 T177 T191 T192 S218 T231 T256 S325 S335 S381 T464 T482 T538 T581 T617 S693 S94 S166 T201 S202 S321 T568 S614 T658 Y459	N7 N49 N462	Leucine zipper: L680-L701	Colon cancer antigen NY-CO-8 g3170180 (Homo sapiens) Scanlan, M.J. et al. (1998) Int. J. Cancer 76:652-658.	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
53	880	S18 S68 T123 T143 S159 T178 T286 S294 S327 S376 S388 T397 T403 S426 S438 S474 S563 T587 T634 T645 S659 S665 S677 S756 S799 S809 T827 S870 S82 T88 S99 T131 T165 S215 S253 S362 S487 T510 S525 S589 T593 S622	M60 N251 N338 N514 N585 N643	Myb1 DNA-binding domain: W808-I816 WD40 domains: L41-N79, K84-N124, T131-D170, G239-D281, A771-S809, F157-T171 Acidic Serine Cluster Repeat: A423-R697	homologous to mouse gene PC326 g458692 (Homo sapiens) Bergsagel, P.L. et al. (1992) Oncogene 7:2059-2064.	BLAST-GenBank BLAST-DOMO HMMER-PFAM BLIMPS-PRINTS MOTIFS
54	855	T460 S8 S179 S261 T288 T313 T377 T706 T719 T755 S764 S803 S851 S34 S67 T129 S190 S339 T391 S483 S502 S537 Y92	N552	Crooked neck protein (RNA processing associated, contains TPR repeat): W398-V814	Predicted TPR domain protein G2315362 (Caenorhabditis elegans) Zhang, K. et al. (1991) Genes Dev. 5:1080-1091.	BLAST-GenBank BLAST-PRODOM MOTIFS



Table 3

Nucleotide Seq ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
55	263-307	Cardiovascular (0.200) Gastrointestinal (0.200) Reproductive (0.200)	Cancer (0.433) Inflammation (0.267) Cell Proliferation (0.200)	PBLUESCRIPT
56	406-450	Reproductive (0.222) Cardiovascular (0.167) Gastrointestinal (0.167) Nervous (0.167)	Cancer (0.500) Inflammation (0.389) Cell Proliferation (0.167)	PSPORT1
57	1001-1045	Reproductive (0.265) Gastrointestinal (0.206) Nervous (0.206)	Cancer (0.412) Inflammation (0.324) Cell Proliferation (0.176)	PINCY
58	226-270	Nervous (0.316) Hematopoietic/Immune (0.211) Reproductive (0.211)	Cancer (0.368) Inflammation (0.368) Cell Proliferation (0.158)	PINCY
59	406-450	Hematopoietic/Immune (0.500) Cardiovascular (0.227)	Cancer (0.182) Inflammation (0.682) Cell Proliferation (0.136)	PINCY
60	56-100	Gastrointestinal (0.545) Nervous (0.182) Reproductive (0.182)	Cancer (0.545) Inflammation (0.364) Cell Proliferation (0.273)	PINCY
61	1046-1090	Nervous (0.271) Reproductive (0.220) Gastrointestinal (0.153)	Cancer (0.542) Inflammation (0.288) Cell Proliferation (0.220)	PINCY
62	226-270	Hematopoietic/Immune (0.288) Nervous (0.178) Reproductive (0.164)	Cancer (0.397) Inflammation (0.548)	PINCY
63	559-603	Reproductive (0.260) Gastrointestinal (0.145) Cardiovascular (0.130)	Cancer (0.458) Inflammation (0.359) Cell Proliferation (0.176)	PSPORT1
64	12-56	Reproductive (0.385) Gastrointestinal (0.231) Cardiovascular (0.154) Nervous (0.154)	Cancer (0.538) Inflammation (0.154) Cell Proliferation (0.154)	PINCY
65	488-532 1091-1135	Reproductive (0.308) Nervous (0.282) Gastrointestinal (0.154)	Cancer (0.487) Inflammation (0.231) Cell Proliferation (0.103)	PINCY





Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
66	37-81	Nervous (0.500) Dermatologic (0.250) Reproductive (0.250)	Inflammation (0.500)	PINCY
67	326-370 1136-1180	Nervous (0.237) Reproductive (0.237) Hematopoietic/Immune (0.158)	Cancer (0.395) Inflammation (0.316) Cell Proliferation (0.158)	PINCY
68	451-495	Nervous (0.312) Reproductive (0.312) Developmental (0.125) Hematopoietic/Immune (0.125) Urologic (0.125)	Cancer (0.562) Inflammation (0.188) Cell Proliferation (0.312)	PINCY
69	64-108	Reproductive (0.233) Nervous (0.174) Cardiovascular (0.140)	Cancer (0.477) Inflammation (0.279) Cell Proliferation (0.198)	PINCY
70	77-121	Cardiovascular (0.500) Musculoskeletal (0.500)	Cancer (0.500) Trauma (0.500)	PBLUESCRIPT
71	164-208	Developmental (0.222) Nervous (0.222)	Cancer (0.444) Cell proliferation (0.222) Trauma (0.222)	PINCY
72	604-648	Reproductive (0.362) Gastrointestinal (0.149) Hematopoietic/Immune (0.128)	Cancer (0.426) Inflammation/Trauma (0.276) Cell proliferation (0.170)	PINCY
73	106-150 1066-1110	Reproductive (0.307) Nervous (0.202) Cardiovascular (0.114)	Cancer (0.482) Inflammation/Trauma (0.307) Cell proliferation (0.175)	PINCY
74	651-695	Hematopoietic/Immune (0.290) Reproductive (0.226) Cardiovascular (0.161)	Inflammation/Trauma (0.451) Cell proliferation (0.230) Cancer (0.320)	PINCY
75	241-285 535-579	Reproductive (0.193) Cardiovascular (0.169) Gastrointestinal (0.157)	Cancer (0.458) Inflammation/Trauma (0.337) Cell proliferation (0.169)	PINCY



Table 3 (cont.)

Nucleotide Seq ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
76	173-217 593-637	Nervous (0.513) Reproductive (0.167)	Inflammation/Trauma (0.371) Cancer (0.333) Cell proliferation (0.141)	PINCY
77	13-57	Reproductive (0.241) Nervous (0.202) Cardiovascular (0.140)	Cancer (0.461) Inflammation (0.180) Cell Proliferation (0.167)	PBLUESCRIPT
78	176-220	Nervous (0.279) Reproductive (0.235) Gastrointestinal (0.147)	Cancer (0.500) Inflammation (0.176) Cell Proliferation (0.162)	PBLUESCRIPT
79	79-123	Nervous (0.280) Cardiovascular (0.160) Developmental (0.160)	Cancer (0.480) Cell Proliferation (0.480) Inflammation (0.160)	PBLUESCRIPT
80	870-914	Nervous (0.571) Reproductive (0.238) Developmental (0.095)	Cancer (0.238) Inflammation (0.381) Cell Proliferation (0.190)	PSPORT1
81	149-194	Nervous (0.216) Reproductive (0.201) Gastrointestinal (0.185)	Cancer (0.432) Inflammation (0.259) Cell Proliferation (0.154)	PINCY
82	150-194	Reproductive (0.375) Cardiovascular (0.125) Endocrine (0.125) Hematopoietic/Immune (0.125) Developmental (0.125) Urologic (0.125)	Cancer (0.375) Inflammation (0.375) Trauma (0.250)	PSPORT1
83	177-221	Reproductive (0.199) Gastrointestinal (0.173) Hematopoietic/Immune (0.128) Nervous (0.128)	Cancer (0.429) Inflammation (0.270) Cell Proliferation (0.186)	PINCY
84	342-386	Reproductive (0.252) Gastrointestinal (0.196) Nervous (0.161)	Cancer (0.483) Inflammation (0.238) Cell Proliferation (0.161)	PINCY
85	124-168	Hematopoietic/Immune (0.308) Cardiovascular (0.154) Nervous (0.154) Gastrointestinal (0.154)	Cancer (0.538) Inflammation (0.308)	PINCY



Table 3 (cont.)

Nucleotide Seq ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
86	238-282	Reproductive (0.277) Cardiovascular (0.181) Nervous (0.169)	Cancer (0.434) Inflammation (0.193) Cell Proliferation (0.157)	PINCY
87	117-161	Reproductive (0.250) Gastrointestinal (0.250) Hematopoietic/Immune (0.115)	Cancer (0.558) Inflammation (0.192) Cell Proliferation (0.115) Trauma (0.115)	PINCY
88	139-183	Nervous (0.237) Reproductive (0.214) Gastrointestinal (0.168)	Cancer (0.397) Inflammation (0.298) Trauma (0.137)	PINCY
89	184-228 352-396	Reproductive (0.556) Nervous (0.222) Hematopoietic/Immune (0.111) Developmental (0.111)	Cancer (0.444) Inflammation (0.333) Cell Proliferation (0.333)	PINCY
90	69-113 879-923	Nervous (0.316) Reproductive (0.193) Hematopoietic/Immune (0.158)	Cancer (0.439) Inflammation (0.211) Cell Proliferation (0.123)	PINCY
91	72-116	Nervous (0.211) Reproductive (0.197) Gastrointestinal (0.158)	Cancer (0.461) Inflammation (0.263) Cell Proliferation (0.211)	PSPORT1
92	489-533	Reproductive (0.274) Nervous (0.217) Gastrointestinal (0.123)	Cancer (0.481) Inflammation (0.189) Cell Proliferation (0.160)	PSPORT1
93	761-805	Reproductive (0.219) Hematopoietic/Immune (0.156) Developmental (0.125)	Cancer (0.312) Cell Proliferation (0.281) Inflammation (0.188) Trauma (0.188)	PSPORT1
94	126-170	Reproductive (0.379) Nervous (0.241) Developmental (0.138)	Cancer (0.414) Cell Proliferation (0.241) Inflammation (0.103)	PBUCSCRIPT
95	1173-1217	Reproductive (0.192) Gastrointestinal (0.192) Nervous (0.173)	Cancer (0.481) Inflammation (0.250) Cell Proliferation (0.212)	PINCY
96	465-509	Hematopoietic/Immune (0.250) Cardiovascular (0.158) Gastrointestinal (0.145)	Inflammation (0.368) Cancer (0.355) Cell Proliferation (0.132)	PINCY



Table 3 (cont.)

Nucleotide Seq ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
97	2427-2471	Nervous (0.224) Reproductive (0.197) Gastrointestinal (0.184)	Cancer (0.474) Cell Proliferation(0.263) Inflammation (0.237)	PINCY
98	23-67	Gastrointestinal (0.270) Reproductive (0.190) Cardiovascular (0.135)	Cancer (0.429) Inflammation (0.278) Cell Proliferation(0.143)	PINCY
99	106-150	Gastrointestinal (0.263) Reproductive (0.263) Nervous (0.158)	Cancer (0.474) Inflammation (0.368) Cell Proliferation(0.211)	PINCY
100	73-117 460-504	Hematopoietic/Immune (0.211) Reproductive (0.211) Cardiovascular (0.105) Developmental (0.105) Gastrointestinal (0.105) Musculoskeletal (0.105)	Cancer (0.474) Inflammation (0.263) Cell Proliferation(0.211)	PSPORT1
101	861-905	Developmental (0.333) Nervous (0.667)	Cell Proliferation(0.333) Trauma (0.333) Neurological (0.333)	PINCY
102	8-52	Developmental (1.000)	Cell Proliferation (1.000)	PINCY
103	199-243	Hematopoietic/Immune (0.143) Nervous (0.179) Reproductive (0.286)	Cancer (0.536) Inflammation (0.250) Cell Proliferation(0.214)	PINCY
104	413-457 908-952	Nervous (0.236) Reproductive (0.222) Gastrointestinal (0.125)	Cancer (0.458) Inflammation (0.236) Cell Proliferation(0.139)	PINCY
105		Reproductive (0.270) Gastrointestinal (0.169) Hematopoietic/Immune 0.101) Developmental (0.101) Nervous (0.101)	Cancer (0.449) Inflammation (0.281) Cell Proliferation(0.258)	PINCY
106	255-299 513-557	Reproductive (0.216) Gastrointestinal (0.196) Nervous (0.157)	Cancer (0.490) Inflammation (0.176) Cell Proliferation(0.176)	PINCY
107	167-211 814-859 1922-1966	Reproductive (0.263) Nervous (0.162) Gastrointestinal (0.141)	Cancer (0.455) Inflammation (0.202) Trauma (0.131)	PINCY





Table 3 (cont.)

Nucleotide Seq ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
108	877-921 2230-2274	Reproductive (0.299) Nervous (0.206) Gastrointestinal (0.134)	Cancer (0.536) Inflammation (0.227) Cell Proliferation(0.124)	PINCY



Table 4

Nucleotide SEQ ID NO:	Library	Library Description
55	KIDNNOT01	Library was constructed using RNA isolated from the kidney tissue of a 64-year-old Caucasian female, who died from an intracranial bleed. Patient history included rheumatoid arthritis.
56	BRSTNOT02	Library was constructed using RNA isolated from diseased breast tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated proliferative fibrocystic changes characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Pathology for the associated tumor tissue indicated an invasive grade 4 mammary adenocarcinoma. Patient history included atrial tachycardia and a benign neoplasm. Family history included cardiovascular and cerebrovascular disease.
57	PLACNOT02	Library was constructed using RNA isolated from the placental tissue of a Hispanic female fetus, who was prematurely delivered at 21 weeks' gestation. Serologies of the mother's blood were positive for CMV (cytomegalovirus).
58	BRAINOT12	Library was constructed using RNA isolated from brain tissue removed from the right frontal lobe of a 5-year-old Caucasian male during a hemispherectomy. Pathology indicated extensive polymicrogyria and mild to moderate gliosis (predominantly subpial and subcortical), which are consistent with chronic seizure disorder. Family history included a cervical neoplasm.
59	SPLNNOT04	Library was constructed using RNA isolated from the spleen tissue of a 2-year-old Hispanic male, who died from cerebral anoxia.
60	LNODNOT03	Library was constructed using RNA isolated from lymph node tissue obtained from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy. On microscopic exam, this tissue was found to be extensively necrotic with 10% viable tumor. Pathology for the associated tumor tissue indicated invasive grade 3-4 squamous cell carcinoma. Patient history included hemangioma. Family history included atherosclerotic coronary artery disease, benign hypertension, congestive heart failure, atherosclerotic coronary artery disease.
61	LIVRTUT01	Library was constructed using RNA isolated from liver tumor tissue removed from a 51-year-old Caucasian female during a hepatic lobectomy. Pathology indicated metastatic grade 3 adenocarcinoma consistent with colon cancer. Family history included a malignant neoplasm of the liver.



Table 4 (cont.)

Nucleotide Seq ID NO:	Library	Library Description
62	BLADTUT07	Library was constructed using RNA isolated from bladder tumor tissue removed from the anterior bladder wall of a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrectomy. Pathology indicated a grade 3 transitional cell carcinoma in the left lateral bladder. Patient history included angina, emphysema, and tobacco use. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
63	LUNGAST01	Library was constructed using RNA isolated from the lung tissue of a 17-year-old Caucasian male, who died from head trauma. Patient history included asthma.
64	LIVRFET02	Library was constructed using RNA isolated from liver tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
65	LUNGNOT23	Library was constructed using RNA isolated from left lobe lung tissue removed from a 58-year-old Caucasian male. Pathology for the associated tumor tissue indicated metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Family history included prostate cancer, breast cancer, and acute leukemia.
66	TESTNOT07	Library was constructed using RNA isolated from testicular tissue removed from a 31-year-old Caucasian male during an unilateral orchiectomy (excision of testis). Pathology indicated a mass containing a large subcapsular hematoma with laceration of the tunica albuginea. The surrounding testicular parenchyma was extensively necrotic.
67	PROSTUT13	Library was constructed using RNA isolated from prostate tumor tissue removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3). Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis. Family history included multiple myeloma, hyperlipidemia, and rheumatoid arthritis
68	LNODNOT11	Library was constructed using RNA isolated from lymph node tissue removed from a 16-month-old Caucasian male who died from head trauma. Patient history included bronchitis.



Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
69	BRSTNOT35	Library was constructed using RNA isolated from breast tissue removed from a 46-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated normal breast parenchyma, bilaterally. The patient presented with hypertrophy of breast and headache. Patient history included obesity, lumbago, glaucoma, and alcohol abuse. Family history included cataract, osteoarthritis, uterine cancer, benign hypertension, hyperlipidemia, alcoholic cirrhosis of the liver, cerebrovascular disease, and type II diabetes.
70	MUSCNOT01	Library was constructed at Stratagene (STR937209), using RNA isolated from the skeletal muscle tissue of a patient with malignant hyperthermia.
71	LUNGNOT14	Library was constructed using RNA isolated from lung tissue removed from the left lower lobe of a 47-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated a grade 4 adenocarcinoma, and the parenchyma showed calcified granuloma. Patient history included benign hypertension and chronic obstructive pulmonary disease. Family history included type II diabetes and acute myocardial infarction.
72	UTRSNOT06	Library was constructed using RNA isolated from myometrial tissue removed from a 50-year-old Caucasian female during a vaginal hysterectomy. Pathology indicated residual atypical complex endometrial hyperplasia. Pathology for the associated tissue removed during dilation and curettage indicated fragments of atypical complex hyperplasia and a single microscopic focus suspicious for grade 1 adenocarcinoma. Patient history included benign breast neoplasm, hypothyroid disease, polypectomy, and arthralgia. Family history included cerebrovascular disease, atherosclerotic coronary artery disease, hyperlipidemia, and chronic hepatitis.
73	PROSTNOT08	Library was constructed using RNA isolated from prostate tumor tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated an adenocarcinoma (Gleason grade 3+4). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst, and hematuria. Family history included tuberculosis, cerebrovascular disease, and arteriosclerotic coronary artery disease.
74	THYMNOT03	Library was constructed using RNA isolated from thymus tissue removed from a 21-year-old Caucasian male during a thymectomy. Pathology indicated an unremarkable thymus and a benign parathyroid adenoma in the right inferior parathyroid. Patient history included atopic dermatitis, a benign neoplasm of the parathyroid, and tobacco use. Family history included atherosclerotic coronary artery disease and benign hypertension.





Table 4 (cont.)

Nucleotide Seq ID NO:	Library	Library Description
75	PENCNOT01	Library was constructed using RNA isolated from penis corpus cavernosum tissue removed from a 53-year-old male. Patient history included untreated penile carcinoma.
76	BRAUNOT01	Library was constructed using RNA isolated from caudate/putamen/nucleus accumbens tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.
77	HUVELPB01	This library was constructed using RNA isolated from HUV-EC-C (ATCC CRL 1730) cells stimulated with cytokine/LPS. RNA was isolated from two pools of HUV-EC-C cells that had been treated with either 4 units/ml TNF-alpha and 2 units/ml gamma IFN for 96 hours, or 1 unit/ml IL-1 beta and 100 ng/ml LPS for 5 hours.
78	HUVENOB01	This library was constructed using RNA isolated from HUV-EC-C (ATCC CRL 1730) cells.
79	HNT2RAT01	This library was constructed at Stratagene (STR937231), using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated with retinoic acid for 24 hours.
80	BRAINOT04	This library was constructed using RNA isolated from the brain tissue of a 44-year-old Caucasian male with a cerebral hemorrhage. The tissue, which contained coagulated blood, came from the choroid plexus of the right anterior temporal lobe. Family history included coronary artery disease and myocardial infarction.
81	BRAITTUT08	This library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 47-year-old Caucasian male during excision of cerebral meningeal tissue. Pathology indicated grade 4 fibrillary astrocytoma with focal tumoral radionecrosis. Patient history included cerebrovascular disease, deficiency anemia, hyperlipidemia and epilepsy. Family history included cerebrovascular disease and a malignant prostate neoplasm.
82	PROSNON01	This library was constructed from 4.4 million independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.



Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
83	PANCTUT01	This library was constructed using RNA isolated from pancreatic tumor tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Previous surgeries included a total splenectomy, cholecystectomy, and abdominal hysterectomy. Family history included cardiovascular disease, type II diabetes, and stomach cancer.
84	BRATUT13	This library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 68-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a meningioma in the left frontal lobe.
85	STOMFET01	This library was constructed using RNA isolated from the stomach tissue of a Caucasian female fetus, who died at 20 weeks' gestation.
86	PROSNOT16	This library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
87	SINTNOT13	This library was constructed using RNA isolated from ileum tissue obtained from a 25-year-old Asian female during a partial colectomy and temporary ileostomy. Pathology indicated moderately active chronic ulcerative colitis, involving colonic mucosa from the distal margin to the ascending colon. Family history included hyperlipidemia, depressive disorder, malignant cervical neoplasm, viral hepatitis A, and depressive disorder.
88	SINTNOT13	This library was constructed using RNA isolated from ileum tissue obtained from a 25-year-old Asian female during a partial colectomy and temporary ileostomy. Pathology indicated moderately active chronic ulcerative colitis, involving colonic mucosa from the distal margin to the ascending colon. Family history included hyperlipidemia, depressive disorder, malignant cervical neoplasm, viral hepatitis A, and depressive disorder.
89	LUNGFET03	This library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
90	SKINBIT01	This library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.



Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
91	LUNGFTUT03	This library was constructed using RNA isolated from lung tumor tissue removed from the left lower lobe of a 69-year-old Caucasian male during segmental lung resection. Pathology indicated residual grade 3 invasive squamous cell carcinoma. Patient history included acute myocardial infarction, prostatic hyperplasia, malignant skin neoplasm, and tobacco use.
92	OVARTUT01	This library was constructed using RNA isolated from ovarian tumor tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology indicated grade 2 mucinous cystadenocarcinoma involving the entire left ovary. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.
93	LUNGFFET05	This library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation from anencephalus.
94	ENDANOT01	This library was constructed using RNA isolated from aortic endothelial cell tissue from an explanted heart removed from a male during a heart transplant.
95	ESOGTUT02	This library was constructed using RNA isolated from esophageal tumor tissue obtained from a 61-year-old Caucasian male during a partial esophagectomy, proximal gastrectomy, pyloromyotomy, and regional lymph node excision. Pathology indicated an invasive grade 3 adenocarcinoma in the esophagus. Family history included atherosclerotic coronary artery disease, type II diabetes, chronic liver disease, primary cardiomyopathy, benign hypertension, and cerebrovascular disease.
96	SINIUCT01	This library was constructed using RNA isolated from ileum tissue obtained from a 42-year-old Caucasian male during a total intra-abdominal colectomy and endoscopic jejunostomy. Previous surgeries included polypectomy, colonoscopy, and spinal canal exploration. Family history included cerebrovascular disease, benign hypertension, atherosclerotic coronary artery disease, and type II diabetes.
97	NPOLNOT01	This library was constructed using RNA isolated from nasal polyp tissue removed from a 78-year-old Caucasian male during a nasal polypectomy. Pathology indicated a nasal polyp and striking eosinophilia. Patient history included asthma and nasal polyps.
98	ADRENOT09	This library was constructed using RNA isolated from left adrenal gland tissue removed from a 43-year-old Caucasian male during nephroureterectomy, regional lymph node excision, and unilateral left adrenalectomy. Pathology for the associated tumor tissue indicated a grade 2 renal cell carcinoma mass in the posterior lower pole of the left kidney with invasion into the renal pelvis.



Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
99	BRAIUNT01	This library was constructed using RNA isolated from SK-N-MC, a neuroepithelioma cell line (ATCC HTB-10) derived from a 14-year-old Caucasian female with neuroepithelioma, with metastasis to the supra-orbital area.
100	LUNGNON03	This library was constructed from 2.56 x 1e6 independent clones from a lung tissue library. RNA was made from lung tissue removed from the left lobe a 58-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated a metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Patient also received radiation therapy to the retroperitoneum. Family history included prostate cancer, breast cancer, and acute leukemia. The normalization and hybridization conditions were adapted from Soares et al., PNAS (1994) 91:9228; Swaroop et al., NAR (1991) 19:1954; and Bonaldo et al., Genome Research (1996) 6:791.
101	BRADDT02	This library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, and emphysema.
102	PLACNOT05	This library was constructed using RNA isolated from placental tissue removed from a Caucasian male fetus, who died after 18 weeks' gestation from fetal demise.
103	HELATX03	This library was constructed using RNA isolated from a treated Hela cell line, derived from cervical adenocarcinoma removed from a 31-year-old Black female. The cells were treated with 1 microm PMA and 100 microm cycloheximide for 24 hours.
104	COLHUT01	This library was constructed using RNA isolated from colon tumor tissue removed from the hepatic flexure of a 55-year-old Caucasian male during right hemicolectomy, incidental appendectomy, and permanent colostomy. Pathology indicated invasive grade 3 adenocarcinoma. Patient history included benign hypertension, anxiety, abnormal blood chemistry, blepharitis, heart block, osteoporosis, acne, and hyperplasia of prostate. Family history included prostate cancer, acute myocardial infarction, stroke, and atherosclerotic coronary artery disease.
105	PLACFER01	This library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus who died after 16 weeks' gestation from fetal demise and hydrocephalus. Serology was positive for CMV antibody.
106	293TF2T01	This library was constructed using RNA isolated from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. The cells were treated with 5-aza-2'-deoxycytidine and transformed with adenovirus 5 DNA.





Table 4 (cont.)

Nucleotide Seq ID NO:	Library	Library Description
107	BRAENOT02	This library was constructed using RNA isolated from posterior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male.
108	FTUBTUT02	This library was constructed using RNA isolated from fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid and serous adenocarcinoma confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma in an endometrial polyp. Metastatic endometrioid and serous adenocarcinoma were present. The patient presented with a pelvic mass and ascites. Patient history included medullary carcinoma of the thyroid and myocardial infarction.



Table 5

Program	Description	Reference	Parameter Threshold
ABIFACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	PE Biosystems, Foster City, CA.	
ABIPARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	PE Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	PE Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLOCKS IMPROVED Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families



Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	



What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group

consisting of:

5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID

NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10,  
SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17,  
SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25,  
SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32,  
SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38,  
SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45,  
SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52,  
SEQ ID NO:53, and SEQ ID NO:54,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an

15 amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID  
NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11,  
SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18,  
SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26,  
SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33,  
SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39,  
SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46,  
SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53,  
and SEQ ID NO:54,

25 c) a biologically active fragment of an amino acid sequence selected from the group

consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID  
NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID  
NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID  
NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID  
NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID  
NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID  
NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID  
NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting  
of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ  
ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID  
NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID





NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, and SEQ ID NO:108.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.



9. A method for producing a polypeptide of claim 1, the method comprising:
- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.
10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, and SEQ ID NO:108,
- b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, and SEQ ID NO:108,

- c) a polynucleotide sequence complementary to a),
- d) a polynucleotide sequence complementary to b), and
- e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.



13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

18. A method for treating a disease or condition associated with decreased expression of functional CCYPR, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of



- claim 1, the method comprising:
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.
20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.
21. A method for treating a disease or condition associated with decreased expression of functional CCYPR, comprising administering to a patient in need of such treatment a composition of claim 20.
22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.
23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.
24. A method for treating a disease or condition associated with overexpression of functional CCYPR, comprising administering to a patient in need of such treatment a composition of claim 23.
25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:
- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.
26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:
- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound





with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and  
b) detecting altered expression of the target polynucleotide.

28. A method for assessing toxicity of a test compound, said method comprising:  
a) treating a biological sample containing nucleic acids with the test compound;  
b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and  
d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.



SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.  
HILLMAN, Jennifer L.  
LAL, Preeti  
TANG, Y. Tom  
YUE, Henry  
AU-YOUNG, Janice  
BANDMAN, Olga  
AZIMZAI, Yalda  
YANG, Junming  
LU, Dying Aina M.  
BAUGHN, Mariah R.  
PATTERSON, Chandra  
SHAH, Purvi

<120> CELL CYCLE AND PROLIFERATION PROTEINS

<130> PF-0722 PCT

<140> To Be Assigned  
<141> Herewith

<150> 60/145,075; 60/153,129; 60/164,647  
<151> 1999-07-21; 1999-09-08; 1999-11-10

<160> 108  
<170> PERL Program

<210> 1  
<211> 145  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc-feature  
<223> Incyte ID No: 116462CD1

<400> 1  
Met Asn Gly Arg Val Asp Tyr Leu Val Thr Gln Gln Ile Asn  
15  
Leu Thr Arg Gly Pro Ser Gly Leu Gly Phe Asn Ile Val Gly Gly  
20  
Thr Asp Gln Gln Tyr Val Ser Asn Asp Ser Gly Ile Tyr Val Ser  
35  
Arg Ile Lys Gln Asn Gly Ala Ala Leu Asp Gly Arg Leu Gln  
50  
Gln Gly Asp Lys Ile Leu Ser Val Asn Gly Gln Asp Leu Lys Asn  
65  
Leu Leu His Gln Asp Ala Val Asp Leu Phe Arg Asn Ala Gly Tyr  
80  
Ala Val Ser Leu Arg Val Gln His Arg Leu Gln Val Gln Asn Gly  
95  
Pro Ile Gly His Arg Gly Gln Gly Asp Pro Ser Gly Ile Pro Ile  
110  
Phe Met Val Leu Val Pro Val Phe Ala Leu Thr Met Val Ala Ala  
125  
Trp Ala Phe Met Arg Tyr Arg Gln Gln Leu  
140  
145

<210> 2  
<211> 340  
<212> PRT



<213> Homo sapiens

<220>

<221> misc-feature

<223> Incyte ID No: 1210462CD1

<400> 2

Met Leu Thr Gln Leu Lys Ala Lys Ser Gln Gly Lys Leu Ala Lys

15

Gln Ile Cys Lys Val Val Leu Asp His Phe Gln Lys Gln Tyr Ser

30

Lys Gln Leu Gly Asp Ala Trp Asn Thr Val Arg Gln Ile Leu Thr

45

Ser Pro Ser Cys Trp Gln Tyr Ala Val Leu Asn Arg Phe Asn

60

Tyr Pro Phe Gln Leu Gln Lys Asp Leu His Leu Lys Gly Tyr His

75

Thr Leu Ser Gln Gly Ser Leu Pro Asn Tyr Pro Lys Ser Val Lys

90

Cys Tyr Leu Ser Arg Thr Pro Gly Arg Ile Pro Ser Gln Arg His

105

Gln Ile Gly Asn Leu Lys Tyr Tyr Leu Asn Ala Ala Ser

120

Leu Leu Pro Val Leu Ala Leu Gln Leu Arg Asp Gly Gln Lys Val

135

Leu Asp Leu Cys Ala Pro Gly Gly Lys Ser Ile Ala Leu Leu

150

Gln Cys Ala Cys Pro Gly Tyr Leu His Cys Asn Gln Tyr Asp Ser

165

Leu Arg Leu Arg Trp Leu Arg Gln Thr Leu Gln Ser Phe Ile Pro

180

Gln Pro Leu Ile Asn Val Ile Lys Val Ser Gln Leu Asp Gly Arg

195

Lys Met Gly Asp Ala Gln Pro Gln Met Phe Asp Lys Val Leu Val

210

Asp Ala Pro Cys Ser Asn Asp Arg Ser Trp Leu Phe Ser Ser Asp

225

Ser Gln Lys Ala Ser Cys Arg Ile Ser Gln Arg Arg Asn Leu Pro

240

Leu Leu Gln Ile Gln Leu Leu Arg Ser Ala Ile Lys Ala Leu Arg

255

Pro Gly Gly Ile Leu Val Tyr Ser Thr Cys Thr Leu Ser Lys Ala

270

Gln Asn Gln Asp Val Ile Ser Gln Ile Leu Asn Ser His Gly Asn

285

Ile Met Pro Met Asp Ile Lys Gly Ile Ala Arg Thr Cys Ser His

300

Asp Phe Thr Phe Ala Pro Thr Gly Gln Cys Gly Leu Leu Val

315

Ile Pro Asp Lys Gly Lys Ala Trp Gly Pro Met Tyr Val Ala Lys

330

Leu Lys Lys Ser Trp Ser Thr Gly Lys Trp

340



Met Leu Tyr Leu Gln Asp Tyr Leu Gln Met Ile Gln Leu Pro  
 15  
 30  
 Met Asp Leu Arg Arg Phe Thr Gln Met Arg Gln Met Asp Leu  
 20  
 25  
 Gln Val Gln Asn Ala Met Asp Gln Leu Gln Arg Val Ser Gln  
 35  
 40  
 Phe phe Met Asn Ala Lys Lys Asn Lys Pro Gln Trp Arg Gln Gln  
 50  
 55  
 Gln Met Ala Ser Ile Lys Lys Asp Tyr Tyr Lys Ala Leu Gln Asp  
 60  
 75  
 Ala Asp Gln Lys Val Gln Leu Ala Asn Gln Ile Tyr Asp Leu Val  
 80  
 85  
 Asp Arg His Leu Arg Lys Leu Asp Gln Gln Leu Ala Lys Phe Lys  
 95  
 100  
 Met Gln Leu Gln Ala Asp Asn Ala Gly Ile Thr Gln Ile Leu Gln  
 110  
 115  
 Arg Arg Ser Leu Gln Leu Asp Thr Pro Ser Gln Pro Val Asn Asn  
 125  
 130  
 His His Ala His Ser His Thr Pro Val Gln Lys Arg Lys Tyr Asn  
 140  
 145  
 Pro Thr Ser His His Thr Thr Thr Asp His Ile Pro Gln Lys Lys  
 155  
 160  
 Phe Lys Ser Gln Ala Leu Leu Ser Thr Leu Thr Ser Asp Ala Ser  
 170  
 175  
 Lys Gln Asn Thr Leu Gln Gly Cys Arg Asn Asn Ser Thr Ala Ser  
 185  
 190  
 Ser Asn Asn Ala Tyr Asn Val Asn Ser Ser Gln Pro Leu Gly Ser  
 200  
 215  
 Tyr Asn Ile Gly Ser Leu Ser Ser Gly Thr Gly Ala Gly Ala Ile  
 225  
 230  
 Thr Met Ala Ala Ala Gln Ala Val Gln Ala Thr Ala Gln Met Lys  
 240  
 255  
 Gln Gly Arg Arg Thr Ser Ser Leu Lys Ala Ser Tyr Gln Ala Phe  
 260  
 275  
 Lys Asn Asn Asp Phe Gln Leu Gly Lys Gln Phe Ser Met Ala Arg  
 285  
 295  
 Thr Gln Asn Ala Ser Ser Ala Ala Asp Ser Arg Ser Gly Arg  
 300  
 315  
 Lys Ser Lys Asn Asn Asn Lys Ser Ser Ser Gln Gln Ser Ser Ser  
 330  
 345  
 Val Val Gln Gln Ile Ser Gln Gln Thr Thr Val Val Pro Gln Ser  
 365  
 375  
 Arg Tyr Cys Ile Cys Asn Gln Val Ser Tyr Gly Gln Met Val Gly  
 380  
 395  
 Cys Asp Asn Gln Asp Cys Pro Ile Gln Trp Phe His Tyr Gly Cys  
 405  
 410  
 Cys Thr Ala Ala Met Lys Arg Arg Gly Ser Arg His Lys

<210> 4  
 <211> 297  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature





&lt;223&gt; Incyte ID No: 1416289CD1

&lt;400&gt; 4

Met Ala Tyr Asn Val Ile Ile Tyr Phe Asn Phe Arg Cys Leu  
1  
5Glu Trp Leu Leu Asn Asn Leu Met Thr His Gln Asn Val Glu Leu  
20  
25Phe Lys Glu Leu Ser Ile Asn Val Met Lys Gln Leu Ile Gly Ser  
35  
40  
45Ser Asn Leu Phe Val Met Gln Val Glu Met Asp Ile Tyr Thr Ala  
50  
55  
60Leu Lys Lys Trp Met Phe Leu Gln Leu Val Pro Ser Trp Asn Gly  
65  
70  
75Ser Leu Lys Gln Leu Thr Glu Thr Asp Val Trp Phe Ser Lys  
80  
85  
90Gln Arg Lys Asp Phe Glu Gly Met Ala Phe Leu Glu Thr Glu Gln  
95  
100  
105Gly Lys Pro Phe Val Ser Val Phe Arg His Leu Arg Leu Gln Tyr  
110  
115  
120Ile Ile Ser Asp Leu Ala Ser Ala Arg Ile Ile Glu Gln Asp Ala  
125  
130  
135Val Val Pro Ser Glu Trp Leu Ser Ser Val Tyr Lys Gln Gln Trp  
140  
145  
150Phe Ala Met Leu Arg Ala Glu Gln Asp Ser Glu Val Gly Pro Gln  
155  
160  
165Glu Ile Asn Lys Glu Leu Glu Gly Asn Ser Met Arg Cys Gly  
170  
175  
180Arg Lys Leu Ala Lys Asp Gly Glu Tyr Cys Trp Arg Trp Thr Gly  
185  
190  
195Phe Asn Phe Gly Phe Asp Leu Leu Val Thr Tyr Thr Asn Arg Tyr  
200  
205  
210Ile Ile Phe Lys Arg Asn Thr Leu Asn Gln Pro Cys Ser Gly Ser  
215  
220  
225Val Ser Leu Gln Pro Arg Arg Ser Ile Ala Phe Arg Leu Arg Leu  
230  
235  
240Ala Ser Phe Asp Ser Ser Gly Lys Leu Ile Cys Ser Arg Thr Thr  
245  
250  
255Gly Tyr Gln Ile Leu Thr Leu Glu Lys Asp Gln Gln Val Val  
260  
265  
270Met Asn Leu Asp Ser Arg Leu Leu Ile Phe Pro Leu Tyr Ile Cys  
275  
280  
285Cys Asn Phe Leu Tyr Ile Ser Pro Glu Lys Asn  
290  
295

&lt;210&gt; 5

&lt;211&gt; 184

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc-feature

&lt;223&gt; Incyte ID No: 1558289CD1

&lt;400&gt; 5

Met Glu Ser Phe Ser Ser Lys Ser Leu Ala Leu Gln Ala Glu Lys  
1  
5  
10  
15Lys Leu Leu Ser Lys Met Ala Gly Arg Ser Val Ala His Leu Phe  
20  
25  
30Ile Asp Glu Thr Ser Ser Glu Val Leu Asp Glu Leu Tyr Arg Val  
35  
40  
45Ser Lys Glu Tyr Thr His Ser Arg Pro Gln Ala Gln Arg Val Ile  
50  
55  
60Lys Asp Leu Ile Lys Val Ala Ile Lys Val Ala Val Leu His Arg  
65  
70  
75



Asn Gly Ser Phe Gly Pro Ser Gln Leu Ala Leu Ala Thr Arg Phe  
 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180  
 Arg Gln Lys Leu Arg Gln Gly Ala Met Thr Ala Leu Ser Phe Gly  
 Thr Val Asp Phe Thr Phe Gln Ala Ala Val Leu Ala Gly Leu Leu  
 Thr Gln Cys Arg Asp Val Leu Leu Gln Leu Val Gln His His Leu  
 Thr Pro Lys Ser His Gly Arg Ile Arg His Val Phe Asp His Phe  
 Ser Asp Pro Gly Leu Thr Ala Leu Tyr Gly Pro Asp Phe Thr  
 Gln His Leu Gly Lys Ile Cys Asp Gly Leu Arg Lys Leu Leu Asp  
 Gln Gly Lys Leu

<210> 6  
 <211> 173  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1577739CD1

<400> 6  
 Met Asp Val Arg Arg Val Leu Val Lys Ala Gln Met Gln Lys Phe  
 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165  
 Thr Leu Gly Phe Gln Ser Pro Gln Arg Val Ser Ala Ala Lys Thr  
 Ala Pro Ile Gln Arg Arg Asp Ile Phe Gln Ser Leu Gln Gly Pro  
 Gln Trp Gln Ser Val Gln Gln Ala Phe Pro His Ile Tyr Ser His  
 Gly Cys Val Leu Lys Asp Val Cys Ser Gln Cys Thr Ser Phe Val  
 Ala Asp Val Val Arg Ser Ser Arg Lys Ser Val Asp Val Leu Asn  
 Thr Thr Pro Arg Arg Ser Arg Gln Thr Gln Ser Leu Tyr Ile Pro  
 Asn Thr Arg Thr Leu Asp Phe Lys

<210> 7  
 <211> 591  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1752768CD1

<400> 7  
 Met Val Pro Val Ala Val Thr Ala Ala Val Ala Pro Val Leu Ser  
 1 5 10 15  
 Ile Asn Ser Asp Phe Ser Asp Leu Arg Gln Ile Lys Lys Gln Leu



20	Leu	Leu	Ile	Ala	Gly	Leu	Thr	Arg	Glu	Arg	Gly	Leu	Leu	His	Ser	30
35	Ser	Lys	Trp	Ser	Ala	Glu	Leu	Ala	Phe	Ser	Leu	Pro	Ala	Leu	Pro	45
50	Leu	Ala	Glu	Leu	Gln	Pro	Pro	Pro	Ile	Thr	Glu	Glu	Asp	Ala	Val	60
65	Leu	Ala	Glu	Leu	Ala	Tyr	Thr	Leu	Ala	Lys	Ala	Tyr	Phe	Asp	Val	75
80	Gln	Asp	Met	Asp	Ala	Tyr	Thr	Leu	Ala	Lys	Ala	Tyr	Phe	Asp	Val	90
95	Lys	Glu	Tyr	Asp	Arg	Ala	Ala	His	Phe	Leu	His	Gly	Cys	Asn	Ser	105
110	Lys	Lys	Ala	Tyr	Phe	Leu	Tyr	Met	Tyr	Ser	Arg	Tyr	Leu	Ser	Gly	120
125	Glu	Lys	Lys	Lys	Asp	Asp	Glu	Thr	Val	Asp	Ser	Leu	Gly	Pro	Leu	135
140	Glu	Lys	Gly	Gln	Val	Lys	Asn	Glu	Ala	Leu	Arg	Glu	Leu	Arg	Val	150
155	Glu	Leu	Ser	Lys	Lys	His	Gln	Ala	Arg	Glu	Leu	Asp	Gly	Phe	Gly	165
170	Leu	Tyr	Leu	Tyr	Gly	Val	Val	Leu	Arg	Lys	Leu	Asp	Leu	Val	Lys	180
185	Glu	Ala	Ile	Asp	Val	Phe	Val	Glu	Ala	Thr	His	Val	Leu	Pro	Leu	195
200	His	Trp	Gly	Ala	Trp	Leu	Glu	Leu	Cys	Asn	Leu	Ile	Thr	Asp	Lys	210
215	Glu	Met	Leu	Lys	Phe	Leu	Ser	Leu	Pro	Asp	Thr	Trp	Met	Lys	Glu	225
230	Phe	Phe	Leu	Ala	His	Ile	Tyr	Thr	Glu	Leu	Gln	Leu	Ile	Glu	Gln	240
245	Ala	Leu	Gln	Lys	Tyr	Gln	Asn	Leu	Ile	Asp	Val	Gly	Phe	Ser	Lys	255
260	Ser	Ser	Tyr	Ile	Val	Ser	Gln	Ile	Ala	Val	Ala	Tyr	His	Asn	Ile	270
275	Arg	Asp	Ile	Asp	Lys	Ala	Leu	Ser	Ile	Phe	Asn	Glu	Leu	Arg	Lys	285
290	Gln	Asp	Pro	Tyr	Arg	Ile	Glu	Asn	Met	Asp	Thr	Phe	Ser	Asn	Leu	300
305	Leu	Tyr	Val	Arg	Ser	Met	Lys	Ser	Glu	Leu	Ser	Tyr	Leu	Ala	His	315
320	Asn	Leu	Cys	Glu	Ile	Asp	Lys	Tyr	Arg	Val	Glu	Thr	Cys	Cys	Val	330
335	Ile	Gly	Asn	Tyr	Tyr	Ser	Leu	Arg	Ser	Gln	His	Glu	Lys	Ala	Ala	345
350	Leu	Tyr	Phe	Gln	Arg	Ala	Leu	Lys	Leu	Asn	Pro	Arg	Tyr	Leu	Gly	360
365	Ala	Trp	Thr	Leu	Met	Gly	His	Glu	Tyr	Met	Glu	Met	Lys	Asn	Thr	375
380	Ser	Ala	Ala	Ile	Gln	Ala	Tyr	Arg	His	Ala	Ile	Glu	Val	Asn	Lys	390
395	Arg	Asp	Tyr	Arg	Ala	Trp	Tyr	Gly	Leu	Gly	Gln	Thr	Tyr	Glu	Ile	405
410	Leu	Lys	Met	Pro	Phe	Tyr	Cys	Leu	Tyr	Tyr	Cys	Arg	Arg	Ala	His	420
425	Gln	Leu	Arg	Pro	Asn	Asp	Ser	Arg	Met	Leu	Val	Ala	Leu	Gly	Glu	435
440	Cys	Tyr	Glu	Lys	Leu	Asn	Gln	Leu	Val	Glu	Ala	Lys	Lys	Cys	Tyr	450
455	Trp	Arg	Ala	Tyr	Ala	Val	Gly	Asp	Val	Glu	Lys	Met	Ala	Leu	Val	465
470	Lys	Leu	Ala	Lys	Leu	His	Glu	Gln	Leu	Thr	Glu	Ser	Glu	Gln	Ala	480
485	Ala	Gln	Cys	Tyr	Ile	Lys	Tyr	Ile	Gln	Asp	Ile	Tyr	Ser	Cys	Gly	495



Glu Ile Val Glu His Leu Glu Glu Ser Thr Ala Phe Arg Tyr Leu  
 500  
 Ala Gln Tyr Tyr Phe Lys Cys Lys Leu Trp Asp Glu Ala Ser Thr  
 510  
 Cys Ala Gln Lys Cys Ala Phe Asn Asp Thr Arg Glu Glu Gly  
 520  
 Lys Ala Leu Leu Arg Gln Ile Leu Gln Leu Arg Asn Gln Gly Glu  
 530  
 Thr Pro Thr Thr Glu Val Pro Ala Pro Phe Leu Pro Ala Ser  
 545  
 Leu Ser Ala Asn Thr Pro Thr Arg Arg Val Ser Pro Leu Asn  
 560  
 Leu Ser Ser Val Thr Pro  
 575  
 585

<210> 8  
 <211> 463  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1887228CD1

<400> 8  
 Met Pro Leu Leu Asn Trp Val Ala Leu Lys Pro Ser Gln Ile Thr  
 1  
 Gly Thr Val Phe Thr Glu Leu Asn Asp Glu Lys Val Leu Gln Glu  
 20  
 Leu Asp Met Ser Asp Phe Glu Glu Gln Lys Thr Lys Ser Gln  
 35  
 Gly Pro Ser Leu Asp Leu Ser Ala Leu Lys Ser Lys Ala Ala Gln  
 50  
 Lys Ala Pro Ser Lys Ala Thr Leu Ile Glu Ala Asn Arg Ala Lys  
 65  
 Asn Leu Ala Ile Thr Leu Arg Lys Gly Asn Leu Gly Ala Glu Arg  
 80  
 Ile Cys Gln Ala Ile Glu Ala Tyr Asp Leu Gln Ala Leu Gly Leu  
 95  
 Asp Phe Leu Glu Leu Leu Met Arg Phe Leu Pro Thr Glu Tyr Glu  
 110  
 Arg Ser Leu Ile Thr Arg Phe Glu Arg Glu Arg Pro Met Glu  
 125  
 Glu Leu Ser Glu Glu Asp Arg Phe Met Leu Cys Phe Ser Arg Ile  
 140  
 Pro Arg Leu Pro Glu Arg Met Thr Thr Leu Thr Phe Leu Gly Asn  
 155  
 Phe Pro Asp Thr Ala Gln Leu Leu Met Pro Gln Leu Asn Ala Ile  
 170  
 Ile Ala Ala Ser Met Ser Ile Lys Ser Ser Asp Lys Leu Arg Gln  
 185  
 Ile Leu Glu Ile Val Leu Ala Phe Gly Asn Tyr Met Asn Ser Ser  
 200  
 Lys Arg Gly Ala Ala Tyr Gly Phe Arg Leu Gln Ser Leu Asp Ala  
 215  
 Leu Leu Glu Met Lys Ser Thr Asp Arg Lys Gln Thr Leu Leu His  
 230  
 Tyr Leu Val Lys Val Ile Ala Glu Lys Tyr Pro Gln Leu Thr Gly  
 245  
 Phe His Ser Asp Leu His Phe Leu Asp Lys Ala Gly Ser Val Ser  
 260  
 Leu Asp Ser Val Leu Ala Asp Val Arg Ser Leu Gln Arg Gly Leu  
 275  
 Glu Leu Thr Gln Arg Glu Phe Val Arg Gln Asp Asp Cys Met Val  
 285





290  
 Leu Lys Gln Phe Arg Ala Asn Ser Pro 295  
 Thr Met Asp Lys Leu 300  
 305  
 Thr Ala Gln Gln Ala Phe Gln Ser Val 310  
 315  
 Leu Ala Asp Ser Lys Thr Ala Gln Gln Ala 320  
 325  
 Gln Tyr Phe Gly Gln Asn Pro Lys Thr 330  
 335  
 Phe Ser Leu Phe Ser Arg Phe Ile Lys 340  
 345  
 Tyr Lys Lys Ala Gln 350  
 355  
 Gln Val Gln Gln Trp Lys Lys Gln Ala 360  
 365  
 Gly Ala Asp Thr Pro Gly Lys Gly Gln Pro 370  
 375  
 Ser 380  
 Pro Pro Lys Ala Arg Pro Gln Met Asp 385  
 Leu Ile Ser Gln Leu 390  
 400  
 Lys Arg Arg Gln Gln Lys Gln Pro Leu Ile 405  
 Tyr Gln Ser Asp Arg 410  
 415  
 Asp Gly Ala Ile Gln Asp Ile Ile Thr Asp 420  
 425  
 Tyr Ile Arg Ala Asp Thr Gly Arg Arg 430  
 435  
 Pro 440  
 445  
 Pro 450  
 Pro Gly Pro Pro Leu 455  
 Gln Val Thr Ser Asp Leu Ser Leu 460

<210> 9  
 <211> 270  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 1988468CD1

<400> 9

Met Ala Asp His Met Met Ala Met Asn His Gly Arg Phe Pro Asp 1  
 5  
 Gly Thr Asn Gly Leu His His His Pro Ala His Arg Met Gly Met 10  
 20  
 Gly Gln Phe Pro Ser Pro His His His Gln Gln Gln Pro Gln 25  
 35  
 His Ala Phe Asn Ala Leu Met Gly Gln His Ile His Tyr Gly Ala 40  
 45  
 Gly Asn Met Asn Ala Thr Ser Gly Ile Arg His Ala Met Gly Pro 50  
 55  
 Gly Thr Val Asn Gly Gly His Pro Ser Ala Leu Ala Pro Ala 60  
 65  
 Ala Arg Phe Asn Asn Ser Gln Phe Met Gly Pro Pro Val Ala Ser 70  
 75  
 Gln Gly Gly Ser Leu Pro Ala Ser Met Gln Leu Gln Lys Leu Asn 80  
 85  
 Asn Gln Tyr Phe Asn His His Pro Tyr Pro His Asn His Tyr Met 90  
 95  
 Pro Asp Leu His Pro Ala Ala Gly His Gln Met Asn Gly Thr Asn 100  
 105  
 Gln His Phe Arg Asp Cys Asn Pro Lys His Ser Gly Gly Ser Ser 110  
 115  
 Thr Pro Gly Gly Ser Gly Gly Ser Ser Thr Pro Gly Gly Ser Gly 120  
 125  
 Ser Ser Ser Ser Gly Gly Ala Gly Ser Ser Asn Ser Gly Gly Gly 130  
 135  
 Ser Gly Ser Gly Ser Val Ala His Val Pro Ala 140  
 145  
 Ala Met Leu Pro 215  
 Pro 200  
 Ser Gly Ser Gly Asn Met Pro Ala Ser Val Ala His Val Pro Ala 205  
 210  
 Ala Met Leu Pro 215  
 Asn Val Ile Asp Thr Asp Phe Ile Asp Gln 220  
 225



Glu Val Leu Met Ser Leu Val Ile Glu Met Gly Leu Asp Arg Ile  
 230 235 240 245 250 255 260 265 270  
 Lys Glu Leu Pro Glu Leu Trp Leu Gly Glu Asn Glu Phe Asp Phe  
 Met Thr Asp Phe Val Cys Lys Glu Glu Pro Ser Arg Val Ser Cys

<210> 10  
 <211> 255  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 2049176CD1

<400> 10  
 Met Val Ser Trp Met Ile Ser Arg Ala Val Val Leu Val Phe Gly  
 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90  
 Leu Leu Ser Pro Tyr Thr Lys Gly Ala Ser Leu Ile Tyr Arg Lys  
 Phe Leu His Pro Leu Ser Ser Lys Glu Arg Glu Ile Asp Asp  
 Tyr Ile Val Glu Ala Lys Glu Arg Gly Tyr Glu Thr Met Val Asn  
 Phe Gly Arg Glu Leu Asn Leu Ala Thr Ala Ala Val Thr  
 Ala Ala Val Lys Ser Glu Gly Ala Ile Thr Glu Arg Leu Arg Ser  
 Phe Ser Met His Asp Leu Thr Thr Ile Glu Gly Asp Glu Pro Val  
 Gly Glu Arg Pro Tyr Glu Pro Leu Pro Ala Lys Lys Lys Ser  
 Lys Pro Ala Pro Ser Glu Ser Ala Gly Tyr Gly Ile Pro Leu Lys  
 Asp Gly Asp Glu Lys Thr Asp Glu Ala Glu Gly Pro Tyr Ser  
 Asp Asn Glu Met Leu Thr His Lys Gly Arg Arg Ser Glu Ser  
 Met Lys Ser Val Lys Thr Thr Lys Gly Arg Lys Glu Val Arg Tyr  
 Gly Ser Leu Lys Tyr Lys Val Lys Lys Arg Pro Glu Val Tyr Phe

<210> 11  
 <211> 533  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 2686765CD1

<400> 11  
 Met Ser Gly Thr Leu Glu Ser Leu Ala Asp Asp Val Ser Ser Met



1	Gly Ser Asp Ser Gln	Ile Asn Gly Leu Ala	10	15
20	Lys Tyr Gly Phe Leu	Gly Gly Ser Gln Tyr	25	30
35	Ser Ser Ile Pro Val	Asp Val Ala Arg Gln	40	45
50	Leu Asp Met Phe Ser	Asn Trp Asp Lys Trp	55	60
65	Gln Lys Val Lys Leu	Arg Cys Arg Lys Gly	70	75
80	Arg Ala Lys Ala Trp	Gln Tyr Leu Ser Asn	85	90
95	Gln Gln Asn Pro Gly	Lys Phe Gln Gln Leu	100	105
110	Asp Pro Lys Trp Leu	Asp Val Ile Gln Lys	115	120
125	Phe Pro Phe His Gln	Met Phe Ala Ala Arg	130	135
140	Gln Asp Leu Tyr Arg	Ile Leu Lys Ala Tyr	145	150
155	Asp Gln Gly Tyr Cys	Gln Ala Gln Ala Val	160	165
170	Leu Met His Met Pro	Ala Gln Lys Pro Phe	175	180
185	Ile Cys Asp Lys Tyr	Leu Pro Gly Tyr Ser	190	195
200	Ala Ile Gln Leu Asp	Gly Gln Ile Phe Ala	205	210
215	Ala Ser Pro Leu Ala	His Arg His Leu Gln	220	225
230	Pro Val Leu Tyr Met	Thr Gln Trp Phe Met	235	240
245	Thr Leu Pro Trp Ala	Ser Val Leu Arg Val	250	255
260	Cys Gln Gly Val Lys	Ile Ile Phe Arg Val	265	270
275	Arg His Thr Leu Gly	Ser Val Gln Lys Leu	280	285
290	Met Tyr Gln Thr Met	Gln Gln Leu Arg Asn	295	300
305	Met Gln Gln Asp Phe	Leu Val His Gln Val	310	315
320	Thr Gln Ala Leu Ile	Gln Arg Gln Asn Ala	325	330
335	Trp Arg Gln Thr Arg	Gly Gln Leu Gln Tyr	340	345
350	Leu His Gly Ser Arg	Ala Ile His Gln Arg	355	360
365	Pro Pro Leu Gly Pro	Ser Ser Ser Leu Ser	370	375
380	Lys Ser Arg Gly Ser	Arg Ala Ala Gly Ser	385	390
395	Pro Pro Val Arg Arg	Pro Ala Pro Gly Pro	400	405
410	Val Thr Ala Gln Gly	Leu His Pro Ser Leu	415	420
425	Asn Ser Thr Pro Leu	Gly Ser Ser Lys Thr	430	435
440	Lys Gln Arg Gln Lys	Arg Gln Lys Gln Lys	445	450
455	Gln Arg Gln Lys Gln	Lys Gln Arg Gln Lys	460	465
470	Gln Arg Gln Lys Gln	Lys Gln Arg Gln Lys	475	480



Gln Gln Lys Gln Arg Gln Lys Gln Gln Lys Gln Arg Gln Lys Gln  
 485 490 495 510 525  
 Gln Lys Lys Lys Ala Gln Arg Lys Lys Lys Lys Lys Lys Lys  
 500 505 515 520  
 Asp Gly Pro Pro Gly Pro His Asp Gly Asp Arg Pro Ser Ala  
 530  
 Gln Ala Arg Gln Asp Ala Tyr Phe

<210> 12  
 <211> 160  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 3215187CD1

<400> 12  
 Met Ala Phe Thr Phe Ala Ala Phe Cys Tyr Met Leu Ser Leu Val  
 1 5 10 15  
 Leu Cys Ala Ala Leu Ile Phe Phe Ala Ile Trp His Ile Ile Ala  
 20 25 30  
 Phe Asp Gln Leu Arg Thr Asp Phe Lys Ser Pro Ile Asp Gln Cys  
 35 40 45  
 Asn Pro Val His Ala Arg Gln Arg Leu Arg Asn Ile Gln Arg Ile  
 50 55 60  
 Cys Phe Leu Leu Arg Lys Leu Val Leu Pro Gln Tyr Ser Ile His  
 65 70 75  
 Ser Leu Phe Cys Ile Met Phe Leu Cys Ala Gln Gln Trp Leu Thr  
 80 85 90  
 Leu Gly Leu Asn Val Pro Leu Leu Phe Tyr His Phe Trp Arg Tyr  
 95 100 105  
 Phe His Cys Pro Ala Asp Ser Ser Gln Leu Ala Tyr Asp Pro Pro  
 110 115 120  
 Val Val Met Asn Ala Asp Thr Leu Ser Tyr Cys Gln Lys Gln Ala  
 125 130 135  
 Trp Cys Lys Leu Ala Phe Tyr Leu Ser Phe Phe Tyr Tyr Leu  
 140 145 150  
 Tyr Cys Met Ile Tyr Thr Leu Val Ser Ser  
 155 160

<210> 13  
 <211> 531  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 3500375CD1

<400> 13  
 Met Ala Asp Val Leu Ser Val Leu Arg Gln Tyr Asn Ile Gln Lys  
 1 5 10 15  
 Lys Gln Ile Val Val Lys Gly Asp Gln Val Ile Phe Gly Gln Phe  
 20 25 30  
 Ser Trp Pro Lys Asn Val Lys Thr Asn Tyr Val Val Trp Gly Thr  
 35 40 45  
 Gly Lys Gln Gly Gln Pro Arg Gln Tyr Tyr Thr Leu Asp Ser Ile  
 50 55 60  
 Leu Phe Leu Leu Asn Val His Leu Ser His Pro Val Tyr Val  
 65 70 75  
 Arg Arg Ala Thr Gln Asn Ile Pro Val Val Arg Arg Pro Asp  
 80 85 90  
 Arg Lys Asp Leu Leu Gly Tyr Leu Asn Gly Gln Ala Ser Thr Ser





95	Ala Ser Ile Asp	Ser Ala Pro Leu	Gln	Ile Gln	Arg	110
110	Ser Thr Gln Val	Lys Arg Ala Ala Asp	Gln	Val	Ala	115
120						120
125						130
135						135
140	Lys Lys Pro Arg	Ile Gln Asp	Gln	Cys	Val	145
145						150
150	Gln Arg Leu Ala	Ala Arg Leu Gln	Gly	Ile	Val	155
155						160
160						165
165	Gln Thr Gln Gln	Ile Arg Ser Leu Ser	Gln	Ala Met Ser	Val	170
170						175
175						180
180	Lys Ile Ala Ala	Ile Lys Ala Lys	Lys	Arg	Ser	185
185						190
190	Thr Ile Lys Thr	Asp Leu Asp Asp	Ile	Thr	Ala Leu Lys	195
200						205
205	Arg Ser Phe Val	Asp Ala Gln Val	Asp	Thr	Arg Asp Ile	210
210						215
215	Ser Arg Gln Arg	Val Trp Arg Thr	Arg	Thr	Ile Leu Gln	220
220						225
225						230
230	Thr Gln Lys Asn	Phe Ser Lys Asn	Ile	Phe	Ala Ile Leu	235
240						245
245	Val Lys Ala Arg	Gln Gln Arg	Pro	Ala	Gln Arg	250
250						255
255	Pro Asn Ala Ala	Pro Val Asp	Pro	Thr	Arg	260
260						265
265	Pro Asn Ala Ala	Pro Thr Asp	Thr	Leu	Arg Thr	270
270						275
275	Ile Pro Ala Ala	Thr Asn Arg	Thr	Asp	Gln Arg	280
280						285
285						290
290	Lys Gln Gln Thr	Gln Gly Phe	Lys	Ile	Thr Met Gly	295
300						305
305	His Gly Met Thr	Leu Lys Ser	Val	Thr	Gln Gly	310
310						315
315						320
320	Lys Thr Gln Thr	Pro Ala Ala Gln	Pro	Val	Pro Val	325
330						335
335	Gln Ala Arg Pro	Pro Asn Gln Lys	Gly	Ser	Arg Thr	340
340						345
345						350
350	Ile Ile Ile Ile	Pro Ala Ala Thr	Thr	Ser	Leu Ile Thr	355
360						365
365	Asn Ala Lys Asp	Leu Gln Asp	Leu	Lys	Phe Val Pro	370
370						375
375						380
380	Gln Lys Lys Lys	Gln Gly Cys	Gln	Arg	Gln Asn Gln	385
390						395
395	Gln Arg Arg Lys	Asp Gln Met	Gln	Pro	Gly Thr Ala Ile	400
400						405
405						410
410	Val Thr Val Pro	Thr Arg Val	Val	Val	Ala Val	415
420						425
425	Pro Gln Asp Thr	Asp Arg Val	Val	Ala	Val Phe	430
430						435
435						440
440	Ala Thr Gln Phe	Lys Gly Trp	Pro	Trp	Leu Pro Asp	445
445						450
450	Pro Val Asp Ile	Phe Ala Lys	Thr	Ala	Ile Ser	455
460						465
465	Pro Val Asp Thr	Lys Trp Asp	Val	Thr	Lys Tyr	470
470						475
475						480
480						485
485	Val Leu Gln Leu	Ser Tyr His	Lys	Arg	His Leu Asp	490
490						495
495						500
500	Phe Leu Arg Phe	Trp Gln Thr	Leu	Asp	Arg Tyr Met	505
510						515
515	Lys Ser His Leu	Arg Phe				520
520						525

<210> 14  
 <211> 165  
 <212> PRT  
 <213> Homo sapiens



<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5080410CD1

<400> 14  
 Met Ala Ser Met Arg Gln Ser Asp Thr Gly Leu Trp Leu His Asn  
 1  
 Lys Leu Gly Ala Thr Asp Gln Leu Trp Ala Pro Pro Ser Ile Ala  
 20  
 Ser Leu Leu Thr Ala Val Ile Asp Asn Ile Arg Leu Cys Phe  
 35  
 His Gly Leu Ser Ser Ala Val Lys Leu Lys Leu Leu Gly Thr  
 50  
 Leu His Leu Pro Arg Arg Thr Val Asp Gln His Pro Ile Leu Pro  
 65  
 Met Lys Gly Ala Leu Met Gln Ile Ile Gln Leu Ala Ser Leu Asp  
 80  
 Ser Asp Pro Trp Val Leu Met Val Ala Asp Ile Leu Lys Ser Phe  
 95  
 Pro Asp Thr Gly Ser Leu Asn Leu Gln Leu Gln Asn Pro  
 110  
 Asn Val Gln Asp Ile Leu Gly Gln Leu Arg Gln Lys Val Gly Gln  
 125  
 Cys Gln Ala Ser Ala Met Leu Pro Leu Gln Cys Gln Tyr Leu Asn  
 140  
 Lys Asn Ala Ala Asp 155  
 155

<210> 15  
 <211> 199  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5218248CD1

<400> 15  
 Met Ser Asn Met Gln Lys His Leu Phe Asn Leu Lys Phe Ala Ala  
 1  
 Lys Gln Leu Ser Arg Ser Ala Lys Lys Cys Asp Lys Gln Lys  
 20  
 Ala Gln Lys Ala Lys Ile Lys Lys Ala Ile Gln Lys Gly Asn Met  
 35  
 Gln Val Ala Arg Ile His Ala Gln Asn Ala Ile Arg Gln Lys Asn  
 50  
 Gln Ala Val Asn Phe Leu Arg Met Ser Ala Arg Val Asp Ala Val  
 65  
 Ala Ala Arg Val Gln Thr Ala Val Thr Met Gly Lys Val Thr Lys  
 80  
 Ser Met Ala Gly Val Val Lys Ser Met Asp Ala Thr Leu Lys Thr  
 95  
 Met Asn Leu Gln Lys Ile Ser Ala Leu Met Asp Lys Phe Gln His  
 110  
 Gln Phe Gln Thr Asp Val Gln Thr Gln Met Gln Met Gln Asp Thr  
 125  
 Met Ser Ser Thr Thr Leu Thr Thr Pro Gln Asn Gln Val Asp  
 140  
 Met Leu Leu Gln Gln Met Ala Asp Gln Ala Gly Leu Asp Leu Asn  
 155  
 Met Gln Leu Pro Gln Gly Gln Thr Gly Ser Val Gly Thr Ser Val  
 170  
 Ala Ser Ala Gln Gln Asp Gln Leu Ser Gln Arg Leu Ala Arg Leu  
 180



185 Arg Asp Gln Val 190 195

<210> 16  
<211> 168  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 058336CD1

<400> 16

Met Ala Phe Asn Asp Cys Phe Ser Leu Asn Tyr Pro Gly Asn Pro 1  
1 Cys Pro Gly Asp Leu Ile Gln Val Phe Arg Pro Gly Tyr Gln His 15  
20 Trp Ala Leu Tyr Leu Gly Asp Gly Tyr Val Ile Asn Ile Ala Pro 30  
35 Val Asp Gly Ile Pro Ala Ser Phe Thr Ser Ala Lys Ser Val Phe 45  
50 Ser Ser Lys Ala Leu Val Lys Met Gln Leu Leu Lys Asp Val Val 60  
65 Gly Asn Asp Thr Tyr Arg Ile Asn Asn Lys Tyr Asp Gln Thr Tyr 75  
80 Pro Pro Leu Pro Val Gln Gln Ile Ile Lys Arg Ser Gln Phe Val 90  
95 Ile Gly Gln Gln Val Ala Tyr Asn Leu Leu Val Asn Asn Cys Gln 105  
110 His Phe Val Thr Leu Leu Arg Tyr Gly Gln Gly Val Ser Gln Gln 120  
125 Ala Asn Arg Ala Ile Ser Thr Val Gln Phe Val Thr Ala Ala Val 135  
140 Gly Val Phe Ser Phe Leu Gly Leu Phe Pro Lys Gly Gln Arg Ala 150  
155 Lys Tyr Tyr 160

Lys Tyr Tyr

<210> 17

<211> 162  
<212> PRT

<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 151148CD1

<400> 17

Met Leu Arg Ala Val Gly Ser Leu Leu Arg Leu Gly Arg Gly Leu 1  
5 Thr Val Arg Cys Gly Pro Gly Ala Pro Leu Gln Ala Thr Arg Arg 15  
20 Pro Ala Pro Ala Leu Pro Pro Arg Gly Leu Pro Cys Tyr Ser Ser 30  
35 Gly Gly Ala Pro Ser Asn Ser Gly Pro Gln Gly His Gly Gln Ile 45  
50 His Arg Val Pro Thr Gln Arg Arg Pro Ser Gln Phe Asp Lys Lys 60  
65 Ile Leu Leu Trp Thr Gly Arg Phe Lys Ser Met Gln Gln Ile Pro 75  
80 Pro Arg Ile Pro Pro Gln Met Ile Asp Thr Ala Arg Asn Lys Ala 105  
95 Arg Val Lys Ala Cys Tyr Ile Met Ile Gly Leu Thr Ile Ile Ala 120



Cys Phe Ala Val Ile Val Ser Ala Lys Arg Ala Val Glu Arg His  
 125 130 135  
 Glu Ser Leu Thr Ser Trp Asn Leu Ala Lys Trp Arg  
 140 145 150  
 Glu Glu Ala Ala Leu Ala Ala Glu Ala Lys Ala Lys  
 155 160  
 <210> 18  
 <211> 246  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc-feature  
 <223> Incyte ID No: 1638819CD1  
 <400> 18  
 Met Ala Gly Tyr Leu Lys Leu Val Cys Val Ser Phe Glu Arg Glu  
 1 5 10  
 Gly Phe His Thr Val Gly Ser Arg Cys Lys Asn Arg Thr Gly Ala  
 20 25  
 Glu His Leu Trp Leu Thr Arg His Leu Arg Asp Pro Phe Val Lys  
 35 40 45  
 Ala Ala Lys Val Glu Ser Tyr Arg Cys Arg Ser Ala Phe Lys Leu  
 50 55  
 Leu Glu Val Asn Glu Arg His Glu Ile Leu Arg Pro Gly Leu Arg  
 65 70 75  
 Val Leu Asp Cys Gly Ala Pro Gly Ala Trp Ser Glu Val Ala  
 80 85  
 Val Glu Lys Val Asn Ala Ala Gly Thr Asp Pro Ser Ser Pro Val  
 95 100  
 Gly Phe Val Leu Gly Val Asp Leu Leu His Ile Phe Pro Leu Glu  
 110 115  
 Gly Ala Thr Phe Leu Cys Pro Ala Asp Val Thr Asp Pro Arg Thr  
 125 130  
 Ser Glu Arg Ile Leu Glu Val Leu Pro Gly Arg Ala Asp Val  
 140 145  
 Ile Leu Ser Asp Met Ala Pro Asn Ala Thr Gly Phe Arg Asp Leu  
 155 160  
 Asp His Asp Arg Leu Ile Ser Leu Cys Leu Thr Leu Leu Ser Val  
 175 180  
 Thr Pro Asp Ile Leu Glu Pro Gly Gly Thr Phe Leu Cys Lys Thr  
 185 190  
 Trp Ala Gly Ser Glu Ser Arg Arg Leu Glu Arg Arg Leu Thr Glu  
 200 205  
 Glu Phe Glu Asn Val Arg Ile Ile Lys Pro Glu Ala Ser Arg Lys  
 215 220  
 Glu Ser Ser Glu Val Tyr Phe Leu Ala Thr Glu Tyr His Gly Arg  
 230 235  
 Lys Gly Thr Val Lys Glu  
 245  
 <210> 19  
 <211> 483  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc-feature  
 <223> Incyte ID No: 1655123CD1  
 <400> 19  
 Met Glu Glu Gly Gly Gly Val Arg Ser Leu Val Pro Gly Gly  
 1 5 10  
 Pro Val Leu Leu Val Leu Cys Gly Leu Leu Glu Ala Ser Gly Gly





20	Gly Arg Ala Leu	Pro	Gln	Leu	Ser	Asp	Asp	25
35	Pro	Gln	Phe	Ser	Leu	Pro	Thr	40
45	Asn	Trp	Pro	Gly	Thr	Gly	Val	55
50	Thr	Gly	Pro	Gln	Phe	Ser	Leu	60
55	Asn	Tyr	Val	Ile	Met	Thr	Ala	70
65	Tyr	Lys	Gln	Asp	Asn	Thr	Ala	75
80	Lys	Tyr	Lys	Cys	Ile	Leu	Pro	85
95	Gln	Gln	Lys	Asp	Tyr	Pro	Arg	100
105	Gln	Gln	Lys	Asp	Pro	Arg	Gln	105
110	Pro	Leu	Phe	Lys	Gln	Ser	Ser	115
125	Trp	Thr	Tyr	Gln	Val	Cys	His	130
135	Gln	Gln	Lys	Gln	Thr	Ile	Arg	135
140	Gln	Gln	Lys	Gln	Thr	His	Gln	145
155	Leu	Gly	Asn	Met	Leu	Ala	Lys	160
170	Gln	Ala	Gln	Lys	Gln	Lys	Ser	175
185	Ile	Gln	Gly	Gln	Met	Thr	Pro	190
200	Gly	Thr	Pro	Cys	Ser	Leu	Lys	205
215	Val	Met	Tyr	Ile	Cys	His	Pro	220
230	Val	Ala	Gln	Val	Thr	Cys	Gln	235
245	Pro	Leu	Leu	Cys	Ser	His	Pro	250
260	Val	Asn	Asp	Ile	Phe	Cys	Gln	265
275	Pro	Leu	Thr	Leu	Arg	Gln	Gln	280
290	Val	Pro	Phe	Arg	Arg	Asn	Lys	295
305	Gln	Gln	Arg	Phe	Pro	Ala	Ile	310
320	Gln	Pro	Val	Leu	Thr	Gly	Thr	325
335	Asp	Asp	Gln	Leu	Ile	Lys	Gln	340
350	Arg	Gly	Gly	Val	Gly	Trp	Lys	355
365	His	Val	His	Gln	Tyr	His	Gln	370
380	Val	Val	Val	Gly	Thr	Trp	Asn	385
395	Lys	Lys	Asn	Thr	Ala	Arg	Ala	400
410	Gln	Thr	Val	Arg	Met	Val	Ser	415
425	Cys	Asp	Ile	Thr	Asp	Lys	Pro	430
440	Cys	Lys	Gln	Ser	Asp	Ser	Pro	445
455	Gln	Pro	His	Ser	Cys	Gln	Tyr	460
470	Ile	Cys	Lys	Ile	Leu	Asp	Thr	475
480	Leu	Pro	Asn					



17/93



18/93



Met Gln Ser Lys Gln Arg Ala Leu Asn Asn Leu Ile Val Gln  
1  
5  
20  
35  
40  
45  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120

Ala Asn Lys Gln Pro Leu Ala Leu Pro Leu Asn Val Ser Gln  
Tyr Cys Val Pro Arg Gln Asn Arg Arg Arg Phe Arg Val Arg Gln  
Pro Ile Leu Gln Tyr Arg Trp Asp Ile Met His Arg Leu Gln  
Pro Gln Ala Arg Met Arg Gln Asn Met Gln Arg Ile Gln  
Gln Val Arg Gln Leu Met Gln Lys Leu Arg Gln Lys Gln Leu Ser  
His Ser Leu Arg Ala Val Ser Thr Asp Pro His His Asp His  
His Asp Gln Phe Cys Leu Met Pro

<210> 23  
<211> 113  
<212> PRT  
<213> Homo sapiens  
<220>  
<221> misc\_feature  
<223> Incyte ID No: 017900CD1  
<400> 23

Met Asp Gln Arg Val Gln Leu Ile Lys Ala Leu Leu Ala Leu Pro  
1  
5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105

Leu Asn Asp Tyr Arg Gln Asp Phe Leu Ala Gln Met Lys Arg Val Phe  
Gln Thr Gln Ser Tyr Met Phe Val Asp Gln Asn Thr Phe Ser Ser  
Asp Ala Leu Lys Val Thr Phe Leu Ile Thr Arg Leu Thr Gln Pro  
Ala Leu Gln Trp Val Ile Pro Tyr Ile Lys Lys Gln Ser Pro Leu  
Leu Asn Asp Tyr Arg Gln Phe Leu Ala Gln Met Lys Arg Val Phe

<210> 24  
<211> 308  
<212> PRT  
<213> Homo sapiens  
<220>  
<221> misc\_feature  
<223> Incyte ID No: 035102CD1  
<400> 24

Met Leu Gln Thr Pro Gln Ser Arg Gln Leu Pro Val Pro Gln Ala  
1  
5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

Ala Leu Ala Ala Lys Ser Leu Ala Arg Arg Ala Tyr Arg Arg  
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln  
Thr Ala Ser Gln Gln Arg Pro Lys Gln Gln Leu Gln Gln Gln Arg  
Gln Gln Gln Lys Asp Gln Gln His Asp Gln Gln Thr Arg Ala Pro





65	Leu	Asn	Arg	Thr	Val	Ala	Glu	Leu	Val	Gln	Phe	Leu	Val	75
80	Leu	Asn	Arg	Thr	Val	Ala	Glu	Leu	Val	Gln	Phe	Leu	Val	90
95	Pro	Ile	Lys	Lys	Val	Met	Glu	Met	Val	Lys	Tyr	Lys	Val	105
110	Val	Ile	Gly	Asp	Leu	Ile	Ala	Arg	Thr	Val	Phe	Gln	Arg	120
110	Val	Ile	Gly	Asp	Leu	Ile	Ala	Arg	Thr	Val	Phe	Gln	Arg	120
125	Ala	Gln	Lys	His	Ser	Thr	Ala	Gln	Ser	Pro	Lys	Gln	Met	135
140	Cys	Tyr	Ser	Asn	Gln	Ser	Asp	Gly	Ser	Ala	Met	Glu	Met	150
1	Met	Phe	Gly	Phe	His	Lys	Pro	Lys	Met	Tyr	Arg	Ser	Ile	Gly
5	Cys	Cys	Ile	Cys	Arg	Ala	Lys	Ser	Ser	Ser	Ser	Ser	Ile	Gly
20	Ser	Lys	Arg	Tyr	Glu	Lys	Asp	Phe	Gln	Ser	Cys	Phe	Gly	Leu
35	Glu	Thr	Arg	Ser	Gly	Asp	Ile	Cys	Asn	Ala	Cys	Val	Leu	Val
50	Lys	Arg	Trp	Lys	Lys	Leu	Pro	Ala	Gly	Ser	Lys	Lys	Asn	Trp
65	His	Val	Val	Asp	Ala	Arg	Ala	Gly	Pro	Ser	Leu	Lys	Thr	Leu
80	Lys	Pro	Lys	Lys	Val	Lys	Thr	Leu	Ser	Gly	Asn	Arg	Ile	Lys
95	Asn	Gln	Ile	Ser	Lys	Leu	Gln	Lys	Glu	Phe	Lys	Arg	His	Asn
110	Asp	Ala	His	Ser	Thr	Ser	Ser	Ser	Ala	Ser	Pro	Ala	Gln	Ser
125	Cys	Tyr	Ser	Asn	Gln	Ser	Asp	Gly	Ser	Asp	Thr	Glu	Met	Ala
140	Cys	Tyr	Ser	Asn	Gln	Ser	Asp	Gly	Ser	Asp	Thr	Glu	Met	Ala



Ser Gly Ser Asn Arg Thr Pro Val Phe Ser Phe Leu Asp Leu Thr  
 155 160 165 170 175 180 185 190 195 210  
 Tyr Trp Lys Arg Gln Lys Ile Cys Cys Gly Ile Ile Tyr Lys Gly  
 Arg Phe Gly Gln Val Leu Ile Asp Thr His Leu Phe Lys Pro Cys  
 Cys Ser Asn Lys Lys Ala Ala Ala Gln Lys Pro Gln Gln Gly  
 200 205 210  
 Pro Gln Pro Leu Pro Ile Ser Thr Gln Trp 220

<210> 26  
 <211> 402  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 926810CD1

<400> 26

Met Ala Ser Ile Ile Ala Arg Val Gly Asn Ser Arg Arg Leu Asn  
 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90  
 Arg Ser Leu Gly Pro Ile Met Ala Ser Met Ala Asp Arg Asn Met  
 Lys Leu Phe Ser Gly Arg Val Val Pro Ala Gln Gly Gln Thr  
 Lys Leu Phe Ser Gly Arg Val Val Met Arg Gln Ala Thr Asn  
 Arg Gly Pro Ala Arg Gln Val Met Arg Val Leu Gln Ala Thr Asn  
 Pro Asn Leu Ser Val Ala Asp Phe Leu Arg Ala Met Lys Leu Val  
 Phe Gly Gln Ser Gln Ser Ser Val Thr Ala His Gly Lys Phe Phe  
 Gly Gly Gln Leu Ser Arg Asp Leu Arg Leu Lys Asp Phe  
 Gly Gly Gln Leu Ser Arg Asp Leu Arg Gln Gln Gln Gln Gln Gln  
 Leu Arg Met Tyr Ala Asn Gln Gln Gln Gln Gln Gln Gln Gln Gln  
 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315  
 Ala Gln Phe Pro Ser Thr Ser Gly Gly Ser Gly Tyr Lys Asn Asn



320  
335  
335  
335  
340  
345  
345  
360  
360  
375  
375  
380  
390  
395  
400  
Ala Pro Gly Gln Tyr Asn Asp Phe Ser Gln Pro Leu

<210> 27  
<211> 93  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1398816CD1

<400> 27  
Met Ser Thr Asp Thr Gly Val Ser Leu Pro Ser Tyr Gln Asp  
1  
5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
Thr Val Gly Ile Ala Gly Phe Ala Ile Val Ala Tyr Gly Leu  
Pro Val Gly Ile Ala Gly Phe Ala Ala Ile Val Ala Tyr Gly Leu  
Tyr Lys Leu Lys Ser Arg Gly Asn Thr Lys Met Ser Ile His Leu  
Ile His Met Arg Val Ala Ala Gln Gly Phe Val Val Gly Ala Met  
Thr Val Gly Met Gly Tyr Ser Met Tyr Arg Gln Phe Trp Ala Lys  
Pro Lys Pro

<210> 28  
<211> 353  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1496820CD1

<400> 28  
Met Asn Arg Gln Asp Arg Asn Val Leu Arg Met Lys Gln Arg Gln  
1  
5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
Ala Ile Pro Lys Pro Thr Val Pro Ser Ala Asp Gln Lys Ser  
Glu Met Lys Asp Phe Ile Gly Asp Arg Ser Ile Pro Lys Leu Val  
Asp Lys Leu Ser Ser Arg Ile Gln Ser Met Leu Gly Asn Tyr Asp  
Ser Ser Pro Leu Phe Ala Gln Pro Tyr Lys Val Thr Ser Lys Gln  
Ser Ser Pro Leu Phe Ala Gln Pro Tyr Lys Val Thr Ser Lys Gln  
Asn Pro Asn Phe Phe Gln Arg His Gly Ser His Gln Ser  
Ser Lys Trp Thr Pro Val Gly Pro Ala Pro Ser Thr Ser Gln Ser  
Gln Lys Arg Ser Ser Gly Leu Gln Ser Gly His Ser Ser Gln Arg



Thr Ser Ala Gly	Ser	Ser	Ser	Gly	Thr	Asn	Ser	Ser	Ser	Ser	His	180
Lys Gly Gln His	Gly	Ser	Gln	His	Ser	Lys	Ser	Arg	Ser	Ser	His	165
His Asp Arg Gln	Ser	Tyr	Asn	Asn	Ser	Gly	Ser	Ser	Ser	Ser	Arg	150
Val Ser Ala Gly	Ser	Ser	Gly	Thr	Asn	Ser	Ser	Ser	Ser	Ser	His	145
Arg Ser Gln Arg	Ser	Arg	Ser	Gly	Thr	Asn	Ser	Ser	Ser	Ser	His	140
His Asp Arg Gln	Ser	Arg	Ser	Arg	Asn	Asn	Ser	Ser	Ser	Ser	His	155
Lys Ser His Gly	Asp	His	His	Ser	Lys	Gln	His	Ser	Ser	Ser	His	170
Pro Gly Lys Pro	Gln	Ala	Val	Ser	Ser	Leu	Asn	Ser	Ser	Ser	His	175
Ser Ser Gln His	Ser	Gln	His	Ser	Lys	Gln	His	Ser	Ser	Ser	His	190
Arg Ser His Gly	Asp	His	His	Ser	Lys	Gln	His	Ser	Ser	Ser	His	205
Lys Ser Pro Arg	Asp	Pro	Asp	Ala	Asn	Trp	Asp	Ser	Pro	Ser	Arg	220
Val Pro Phe Ser	Ser	Gly	Gln	His	Ser	Thr	Gln	Ser	Phe	Pro	Pro	235
Ser Leu Met Ser	Lys	Ser	Asn	Ser	Met	Leu	Gln	Lys	Pro	Thr	Ala	250
Tyr Val Arg Pro	Met	Asp	Gly	Gln	Gln	Ser	Met	Gln	Pro	Lys	Leu	265
Ser Ser Gln His	Ser	Ser	Ser	Gln	Ser	His	Gly	Asn	Ser	Met	Thr	280
Gln Leu Lys Pro	Ser	Lys	Ala	His	Leu	Ser	Thr	Lys	Leu	Lys	Ile	295
Pro Ser Gln Pro	Leu	Asp	Ala	Ser	Ala	Ser	Gly	Asp	Val	Ser	Cys	310
Val Asp Gln Ile	Leu	Lys	Gln	Met	Thr	His	Ser	Trp	Pro	Pro	Pro	325
Leu Thr Ala Ile	His	Thr	Pro	Cys	Lys	Thr	Gln	Pro	Ser	Lys	Phe	340
Pro Phe Pro Thr	Lys	Val	Ser	Lys								350
<210> 29												
<211> 120												
<212> PRT												
<213> Homo sapiens												
<220>												
<221> misc-feature												
<223> Incyte ID No:	1514559CD1											
<400> 29												
Met Ser Gln Pro	Ala	Gly	Asp	Val	Arg	Gln	Asn	Pro	Cys	Gly	Ser	15
Lys Ala Cys Arg	Arg	Leu	Phe	Gly	Pro	Val	Asp	Ser	Gln	Gln	Leu	30
Ser Arg Asp Cys	Asp	Ala	Leu	Met	Ala	Gly	Cys	Ile	Gln	Gln	Ala	45
Arg Gln Arg	Trp	Asn	Phe	Asp	Phe	Val	Thr	Gln	Thr	Pro	Leu	60
Gly Asp Phe Ala	Trp	Gln	Arg	Val	Arg	Gly	Leu	Gly	Leu	Pro	Lys	75
Leu Tyr Leu	Pro	Thr	Trp	Ser	Ala	Gly	Tyr	Pro	Leu	Gln	Gly	90
Cys Gly Ser	Phe	Pro	Ser	Leu	Ser	Gln	Ala	Val	Met	Lys	Phe	105
Pro Phe Pro	His	Ser	Asp	Leu	Asn	Ser	Phe	Ser	Phe	Gln	Lys	120
<210> 30												
<211> 144												
<212> PRT												
<213> Homo sapiens												
<220>												





<221> misc-feature  
<223> Incyte ID No: 1620092CD1

<400> 30  
Met Arg Ser Cys Phe Arg Leu Cys Glu Arg Asp Val Ser Ser  
1  
Leu Arg Leu Thr Arg Ser Ser Asp Leu Lys Arg Ile Asn Gly Phe  
20  
Cys Thr Lys Pro Gln Ser Pro Gly Ala Pro Ser Arg Thr Tyr  
35  
Asn Arg Val Pro Leu His Lys Pro Thr Asp Trp Gln Lys Ile  
50  
Leu Ile Trp Ser Gly Arg Phe Lys Lys Glu Asp Gln Ile Pro Gln  
65  
Thr Val Ser Leu Glu Met Leu Asp Ala Ala Lys Asn Lys Met Arg  
80  
Val Lys Ile Ser Tyr Leu Met Ile Ala Leu Thr Val Val Gly Cys  
95  
Ile Phe Met Val Ile Glu Gly Lys Lys Ala Ala Gln Arg His Glu  
110  
Thr Leu Thr Ser Leu Asn Leu Glu Lys Lys Ala Arg Leu Lys Glu  
125  
Glu Ala Ala Met Lys Ala Lys Thr Glu  
140

<210> 31  
<211> 933  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc-feature  
<223> Incyte ID No: 1678765CD1

<400> 31  
Met Phe Tyr Leu Glu Asp Asp Lys Glu Asp Glu Val Val Cys Lys  
1  
Gly Ser Leu Ser Lys Thr Gln Asp Val Tyr His Asp Lys Ser Pro  
20  
Pro Gly Ile Leu Ser Gln Thr Met Asn Tyr Val Gly Gln Leu Ala  
35  
Gly Gln Val Ile Val Thr Val Lys Glu Leu Tyr Lys Gly Ile Asn  
50  
Gln Ala Thr Leu Ser Gly Cys Ile Asp Val Ile Val Val Gln Gln  
65  
Gln Asp Gly Ser Tyr Gln Cys Ser Pro Phe His Val Arg Phe Gly  
80  
Lys Leu Gly Val Leu Arg Ser Lys Glu Lys Val Ile Asp Ile Glu  
95  
Ile Asn Gly Ser Ala Val Asp Leu His Met Lys Leu Gly Asp Asn  
110  
Gly Gln Ala Phe Val Gln Glu Thr Glu Tyr Glu Lys  
125  
Leu Pro Ala Tyr Leu Ala Thr Ser Pro Ile Pro Thr Glu Asp Gln  
140  
Phe Phe Lys Asp Ile Asp Thr Pro Leu Val Lys Ser Gly Gly Asp  
155  
Glu Thr Pro Ser Ser Ser Ser Asp Ile Ser His Val Leu Glu Thr  
170  
Glu Thr Ile Phe Thr Pro Ser Ser Val Lys Lys Lys Arg Arg  
185  
Arg Lys Lys Tyr Lys Glu Asp Ser Lys Lys Glu Gln Ala Ala  
200  
Ser Ala Ala Ala Glu Asp Thr Cys Asp Val Gly Val Ser Ser Asp  
210



225	Ser	Asn	Ala	Ser	220
240	Gly	Leu	Pro	Ala	235
255	Ser	His	Pro	Ala	250
270	Thr	Leu	Pro	Ala	265
285	Glu	Leu	Pro	Ala	280
300	Glu	Leu	Pro	Ala	295
315	Arg	Leu	Pro	Ala	310
330	Ser	Leu	Pro	Ala	325
345	Leu	Leu	Pro	Ala	340
360	Thr	Leu	Pro	Ala	355
375	Thr	Leu	Pro	Ala	370
390	Thr	Leu	Pro	Ala	385
405	Ala	Leu	Pro	Ala	400
420	Val	Leu	Pro	Ala	415
435	His	Leu	Pro	Ala	430
450	Pro	Leu	Pro	Ala	445
465	Gly	Leu	Pro	Ala	460
480	Ser	Leu	Pro	Ala	475
495	Cys	Leu	Pro	Ala	490
510	Leu	Leu	Pro	Ala	505
525	Pro	Leu	Pro	Ala	520
540	Gly	Leu	Pro	Ala	535
555	Tyr	Leu	Pro	Ala	550
570	Val	Leu	Pro	Ala	565
585	Lys	Leu	Pro	Ala	580
600	Lys	Leu	Pro	Ala	595
615	Gly	Leu	Pro	Ala	610
630	Glu	Leu	Pro	Ala	625
645	Asp	Leu	Pro	Ala	640
660	Pro	Leu	Pro	Ala	655
675	Ser	Leu	Pro	Ala	670
690	Asp	Leu	Pro	Ala	685
215	Ala	Leu	Pro	Ala	230
230	Gly	Leu	Pro	Ala	245
245	Glu	Leu	Pro	Ala	260
260	Leu	Leu	Pro	Ala	275
275	Pro	Leu	Pro	Ala	290
290	Ala	Leu	Pro	Ala	305
305	Thr	Leu	Pro	Ala	320
320	Arg	Leu	Pro	Ala	335
335	His	Leu	Pro	Ala	350
350	Pro	Leu	Pro	Ala	365
365	Val	Leu	Pro	Ala	380
380	Thr	Leu	Pro	Ala	395
395	Ser	Leu	Pro	Ala	410
410	Leu	Leu	Pro	Ala	425
425	Pro	Leu	Pro	Ala	440
440	Ala	Leu	Pro	Ala	455
455	Val	Leu	Pro	Ala	470
470	Gly	Leu	Pro	Ala	485
485	Ser	Leu	Pro	Ala	500
500	Met	Leu	Pro	Ala	515
515	Leu	Leu	Pro	Ala	530
530	Pro	Leu	Pro	Ala	545
545	Asn	Leu	Pro	Ala	560
560	Tyr	Leu	Pro	Ala	575
575	Pro	Leu	Pro	Ala	590
590	Met	Leu	Pro	Ala	605
605	Thr	Leu	Pro	Ala	620
620	Pro	Leu	Pro	Ala	635
635	Arg	Leu	Pro	Ala	650
650	Glu	Leu	Pro	Ala	665
665	Ser	Leu	Pro	Ala	680
680	Leu	Leu	Pro	Ala	695



Gly Pro Asn Asp Val Phe Ser Ile Thr Gln Tyr Gln Gly  
 695  
 Thr Cys Arg Cys Ala Gly Thr Ile Tyr Leu Trp Asn Asp  
 705  
 Lys Ile Ile Ile Ser Asp Ile Asp Gly Thr Lys Ser Asp  
 720  
 Ala Leu Gly Gln Ile Leu Pro Gln Leu Gly Lys Asp Trp Thr His  
 735  
 Gln Gly Ile Ala Lys Leu Tyr His Ser Ile Asn Gln Asn Gly Tyr  
 750  
 Lys Phe Leu Tyr Cys Ser Ala Arg Ala Ile Gly Met Ala Asp Met  
 765  
 Thr Arg Gly Tyr Thr His Trp Val Asn Asp Lys Gly Thr Ile Leu  
 780  
 Pro Arg Gly Pro Leu Met Leu Ser Pro Ser Ser Leu Phe Ser Ala  
 795  
 Phe His Arg Gln Val Ile Gln Lys Lys Phe Lys Ile  
 810  
 Gln Cys Leu Asn Asp Ile Lys Asn Leu Phe Ala Pro Ser Lys Gln  
 825  
 Pro Phe Tyr Ala Ala Phe Gly Asn Arg Pro Asn Asp Val Tyr Ala  
 840  
 Tyr Thr Gln Val Gly Val Pro Asp Cys Arg Ile Phe Thr Val Asn  
 855  
 Ser Lys Gly Gln Leu Ile Gln Gln Arg Thr Lys Gly Asn Lys Ser  
 870  
 Ser Tyr His Arg Leu Ser Gln Leu Val Gln His Val Phe Pro Leu  
 885  
 Leu Ser Lys Gln Asn Ser Ala Phe Pro Cys Pro Gln Phe Ser  
 900  
 Ser Phe Cys Tyr Trp Arg Asp Pro Ile Pro Gln Val Asp Leu Asp  
 915  
 Asp Leu Ser  
 930

<210> 32  
 <211> 268  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 1708229CD1

<400> 32  
 Met Leu Gly Asp His Cys Ser Leu Pro Gln Asp Gln Ala Arg Pro  
 1  
 Gly Gln Ser Leu Gln Ser Gly Leu Cys Lys Met Val Leu Gln  
 20  
 Ala Val Ser Lys Val Leu Arg Lys Ser Lys Ala Lys Pro Asn Gly  
 35  
 Lys Lys Pro Ala Ala Gln Gln Arg Lys Ala Tyr Leu Gln Pro Gln  
 45  
 His Thr Lys Ala Arg Ile Thr Asp Phe Gln Phe Lys Gln Leu Val  
 50  
 Val Leu Pro Arg Gln Ile Asp Leu Asn Gln Trp Leu Ala Ser Asn  
 65  
 Thr Thr Thr Phe Phe His His Ile Asn Leu Gln Tyr Ser Thr Ile  
 80  
 Ser Gln Phe Cys Thr Gly Gln Thr Cys Gln Thr Met Ala Val Cys  
 95  
 Asn Thr Gln Tyr Trp Tyr Asp Gln Arg Gly Lys Lys Val Lys  
 110  
 Cys Thr Ala Pro Gln Tyr Val Asp Phe Val Met Ser Ser Val Gln  
 125  
 Cys Thr Ala Pro Gln Tyr Val Asp Phe Val Met Ser Ser Val Gln  
 135



```
<210> 33  
<211> 337  
<212> PRT  
<213> Homo sapiens
```

Met Leu Leu Gly Leu	5	Ala	Ala	Met	Glu	Leu	Lys	Val	Trp	Val	Asp	15	1	Gly	Ile	Gln	Arg	Val	20	Ala	Leu	Ala	Gln	Ala	Ile	Val	Val	Val	35	Arg	phe	Val	Leu	Val	50	Leu	Pro	Gln	Glu	Cys	Pro	Val	60	Gln	Thr	Cys	Gly	75	Val	Ala	Arg	Val	80	Leu	Ala	Gly	Arg	Pro	95	Cys	Leu	Ile	Arg	Ala	110	Gly	Cys	Glu	Pro	Arg	Lys	Thr	Leu	Thr	Pro	115	Leu	Ala	120	Ser	Pro	Ala	Pro	Ser	135	Leu	Ser	Arg	Pro	140	Cys	Cys	Thr	Asp	Leu	Arg	Gly	Leu	Glu	Leu	Arg	Val	Gln	Arg	Asn	165	Ala	Glu	Glu	Leu	Gly	His	Glu	Ala	phe	Trp	Glu	Gln	Glu	Leu	Arg	180	Arg	Glu	Gln	Ala	Arg	Glu	Gly	Gln	Ala	Arg	Leu	Gln	Ala	195	Leu	Ser	Ala	Ala	Thr	Ala	Ala	Arg	Leu	Gln	Ala	Leu	210	Asp	Ala	Gln	Ala	Arg	Ala	Glu	Leu	Gln	Leu	Ala	Ala	225	Glu	Ala	Pro	Gly	Pro	Ser	Pro	Met	Ala	Ser	Ala	Thr	Glu	Arg	240	Val	His	Gln	Asp	Leu	Ala	Val	Gln	Glu	Arg	250	Gln	Gly	Ser	Leu	Ala	260	Met	Leu	Leu	Gly	Leu	1	<400> 33
---------------------	---	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	---	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---	----------





Arg Ala Leu Gln Ala Gln Ala Gln Gln Leu Gln Gln Leu Asn Arg  
 275 280 285  
 Gln Leu Arg Gln Cys Asn Leu Gln Gln Phe Ile Gln Gln Thr Gly  
 290 295 300  
 Ala Ala Leu Pro Pro Pro Arg Gly Pro Gly  
 305 310 315  
 Thr Gln Val Gly Val Leu Gly Gly Gly Trp Gln Val Arg Thr  
 320 325 330  
 Trp Pro Ser Pro Thr Pro Ser

<210> 34  
 <211> 565  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 1806850CD1

<400> 34  
 Met Lys Gln Gln Gln Val Phe Gln Pro Met Leu Met Gln Tyr  
 1 5 10  
 Phe Thr Tyr Gln Gln Leu Lys Tyr Ile Lys Lys Val Ile Ala  
 20 25 30  
 Gln His Cys Ser Gln Lys Asp Thr Ala Gln Leu Leu Arg Gly Leu  
 35 40 45  
 Ser Leu Trp Asn His Ala Gln Gln Arg Gln Lys Phe Phe Lys Tyr  
 50 55 60  
 Ser Val Asp Gln Lys Ser Asp Lys Gln Ala Gln Val Ser Gln His  
 65 70 75  
 Ser Thr Gly Ile Thr His Leu Pro Pro Gln Val Met Leu Ser Ile  
 80 85 90  
 Phe Ser Tyr Leu Asn Pro Gln Gln Leu Cys Arg Cys Ser Gln Val  
 95 100 105  
 Ser Met Lys Trp Ser Gln Leu Thr Lys Thr Gly Ser Leu Trp Lys  
 110 115 120  
 His Leu Tyr Pro Val His Trp Ala Arg Gly Asp Trp Tyr Ser Gly  
 125 130 135  
 Pro Ala Thr Gln Leu Asp Thr Gln Pro Asp Gln Trp Val Lys  
 140 145 150  
 Asn Arg Lys Asp Gln Ser Arg Ala Phe His Gln Trp Asp Gln Asp  
 155 160 165  
 Ala Asp Ile Asp Gln Ser Gln Gln Ser Ile Ala  
 170 175 180  
 Ile Ser Ile Ala Gln Met Gln Lys Arg Leu Leu His Gly Leu Ile  
 185 190 195  
 His Asn Val Leu Pro Tyr Val Gly Thr Ser Val Lys Thr Leu Val  
 200 205 210  
 Leu Ala Tyr Ser Ser Ala Val Ser Ser Lys Met Val Arg Gln Ile  
 215 220 225  
 Leu Gln Leu Cys Pro Asn Leu Gln His Leu Asp Leu Thr Gln Thr  
 230 235 240  
 Asp Ile Ser Asp Ser Ala Phe Asp Ser Trp Ser Trp Leu Gly Cys  
 245 250 255  
 Cys Gln Ser Leu Arg His Leu Asp Leu Ser Gly Cys Gln Lys Ile  
 260 265 270  
 Thr Asp Val Ala Leu Gln Lys Ile Ser Arg Ala Leu Gly Ile Leu  
 275 280 285  
 Thr Ser His Gln Ser Gly Phe Leu Lys Thr Ser Thr Ser Lys Ile  
 290 295 300  
 Thr Ser Thr Ala Trp Lys Asn Lys Asp Ile Thr Met Gln Ser Thr  
 305 310 315  
 Lys Gln Tyr Ala Cys Leu His Asp Leu Thr Asn Lys Gly Ile Gly

.

.

.

.

320	Glu	Asn	Glu	Ile	Asp	325	Thr	Lys	Pro	Val	Ser	330
335	Glu	Asn	Glu	Ile	Asp	340	Thr	Lys	Pro	Val	Ser	345
350	Glu	Asn	Glu	Ile	Asp	355	Thr	Lys	Pro	Val	Ser	360
365	Glu	Asn	Glu	Ile	Asp	370	Thr	Lys	Pro	Val	Ser	375
380	Glu	Asn	Glu	Ile	Asp	385	Thr	Lys	Pro	Val	Ser	390
395	Glu	Asn	Glu	Ile	Asp	400	Thr	Lys	Pro	Val	Ser	405
410	Glu	Asn	Glu	Ile	Asp	415	Thr	Lys	Pro	Val	Ser	420
425	Glu	Asn	Glu	Ile	Asp	430	Thr	Lys	Pro	Val	Ser	435
440	Glu	Asn	Glu	Ile	Asp	445	Thr	Lys	Pro	Val	Ser	450
455	Glu	Asn	Glu	Ile	Asp	460	Thr	Lys	Pro	Val	Ser	465
470	Glu	Asn	Glu	Ile	Asp	475	Thr	Lys	Pro	Val	Ser	480
485	Glu	Asn	Glu	Ile	Asp	490	Thr	Lys	Pro	Val	Ser	495
500	Glu	Asn	Glu	Ile	Asp	505	Thr	Lys	Pro	Val	Ser	510
515	Glu	Asn	Glu	Ile	Asp	520	Thr	Lys	Pro	Val	Ser	525
530	Glu	Asn	Glu	Ile	Asp	535	Thr	Lys	Pro	Val	Ser	540
545	Glu	Asn	Glu	Ile	Asp	550	Thr	Lys	Pro	Val	Ser	555
560	Glu	Asn	Glu	Ile	Asp	565	Thr	Lys	Pro	Val	Ser	570
5	Ser	Phe	Met	Gln	Gly	10	Ser	Ala	Asn	Ile	Arg	15
20	Ser	Phe	Met	Gln	Gly	25	Ser	Ala	Asn	Ile	Arg	30
35	Ser	Phe	Met	Gln	Gly	40	Ser	Ala	Asn	Ile	Arg	45
50	Ser	Phe	Met	Gln	Gly	55	Ser	Ala	Asn	Ile	Arg	60
65	Ser	Phe	Met	Gln	Gly	70	Ser	Ala	Asn	Ile	Arg	75
80	Ser	Phe	Met	Gln	Gly	85	Ser	Ala	Asn	Ile	Arg	90
95	Ser	Phe	Met	Gln	Gly	100	Ser	Ala	Asn	Ile	Arg	105
110	Ser	Phe	Met	Gln	Gly	115	Ser	Ala	Asn	Ile	Arg	120
125	Ser	Phe	Met	Gln	Gly	130	Ser	Ala	Asn	Ile	Arg	135
140	Ser	Phe	Met	Gln	Gly	145	Ser	Ala	Asn	Ile	Arg	150

<210> 35  
 <211> 228  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc-feature  
 <223> Incyte ID No: 1851534CD1



Pro Gln Ile Lys Lys Ile Thr Lys Thr Ile Gln Asn Gly Arg  
155  
160  
165  
Glu Leu Phe Gln Ser  
170  
175  
180  
Gln Ala Ser Gln His Thr Lys Ser Lys His Gln Ser Arg Lys Gln  
185  
190  
195  
Lys Arg Lys Lys Ser Asn Lys His Asp Ser Ser Arg Ser Gln Gln  
200  
205  
210  
Arg Lys Ser His Lys Ile Pro Lys Leu Gln Pro Gln Gln Asn  
215  
220  
225  
Met Thr Lys

<210> 36  
<211> 495  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc-feature  
<223> Incyte ID No: 1868749CD1

<400> 36

Met Lys Gly Met Lys Val Gln Val Leu Asn Ser Asp Ala Val Leu  
1  
5  
10  
Val Ile Ala Ser Val Ile Gln Thr Ala Gly  
15  
20  
25  
Tyr Arg Val Leu Leu Arg Tyr Gln Gly Phe Gln Asn Asp Ala Ser  
30  
35  
36  
His Asp Phe Trp Cys Asn Leu Gly Thr Val Asp Val His Pro Ile  
40  
45  
50  
Gly Trp Cys Ala Ile Asn Ser Lys Ile Leu Val Pro Pro Arg Thr  
55  
60  
65  
Ile His Ala Lys Phe Thr Asp Trp Lys Gly Tyr Leu Met Lys Arg  
70  
75  
80  
Leu Val Gly Ser Arg Thr Leu Pro Val Asp Phe His Ile Lys Met  
85  
90  
95  
Val Gln Ser Met Lys Tyr Pro Phe Arg Gln Gly Met Arg Leu Gln  
100  
105  
110  
Val Val Asp Lys Ser Gln Val Ser Arg Thr Arg Met Ala Val Val  
115  
120  
125  
Asp Thr Val Ile Gly Arg Leu Arg Leu Tyr Gln Asp Gly  
130  
135  
140  
Asp Ser Asp Asp Phe Trp Cys His Met Trp Ser Pro Leu Ile  
145  
150  
155  
His Pro Val Gly Trp Ser Arg Arg Val Gly His Gly Ile Lys Met  
160  
165  
170  
Ser Gln Arg Arg Asp Met Ala His His Pro Thr Phe Arg Lys  
175  
180  
185  
Ile Tyr Cys Asp Ala Val Pro Tyr Leu Phe Lys Lys Val Arg Ala  
190  
195  
200  
Val Tyr Thr Gln Gly Gly Trp Phe Gln Gly Met Lys Leu Gln  
205  
210  
215  
Ala Ile Asp Pro Leu Asn Leu Gly Asn Ile Cys Val Ala Thr Val  
220  
225  
230  
Cys Lys Val Leu Leu Asp Gly Tyr Leu Met Ile Cys Val Asp Gly  
235  
240  
245  
Gly Pro Ser Thr Asp Gly Leu Asp Trp Phe Cys Tyr His Ala Ser  
250  
255  
260  
Ser His Ala Ile Phe Pro Ala Thr Phe Cys Gln Lys Asn Asp Ile  
265  
270  
275  
Gln Leu Thr Pro Pro Lys Gly Tyr Gln Ala Gln Thr Phe Asn Trp  
280  
285  
290  
Gln Asn Tyr Leu Gln Lys Thr Lys Ser Lys Ala Ala Pro Ser Arg  
295  
300



305  
 310  
 315  
 320  
 325  
 330  
 335  
 340  
 345  
 350  
 355  
 360  
 365  
 370  
 375  
 380  
 385  
 390  
 395  
 400  
 405  
 410  
 415  
 420  
 425  
 430  
 435  
 440  
 445  
 450  
 455  
 460  
 465  
 470  
 475  
 480  
 485  
 490  
 495

Leu Phe Asn Met Asp Cys Pro Asn His Gly Phe Lys Val Gly Met  
 Lys Leu Gln Ala Val Asp Leu Met Gln Pro Arg Leu Ile Cys Val  
 Ala Thr Val Lys Arg Val Val His Arg Leu Ser Ile His Phe  
 Asp Gly Trp Asp Ser Gln Tyr Asp Gln Trp Val Asp Cys Gln Ser  
 Pro Asp Ile Tyr Pro Val Gly Trp Cys Gln Leu Thr Gly Tyr Gln  
 Leu Gln Pro Pro Val Ala Ala Gln Pro Ala Thr Pro Leu Lys Ala  
 Lys Gln Ala Thr Lys Lys Lys Lys Gln Phe Gly Lys Lys Arg  
 Lys Arg Ile Pro Thr Lys Thr Arg Pro Leu Arg Gln Gly Ser  
 Lys Lys Pro Leu Gln Asp Asp Pro Gln Gly Ala Arg Lys Ile  
 Ser Ser Gln Pro Val Pro Gly Gln Ile Ala Val Arg Val Lys  
 Gln Gln His Leu Asp Val Ala Ser Pro Asp Lys Ala Ser Ser Pro  
 Gln Leu Pro Val Ser Val Gln Asn Val Tyr

<210> 37  
 <211> 1336  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 1980010CD1

<400> 37  
 1  
 5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65  
 70  
 75  
 80  
 85  
 90  
 95  
 100  
 105  
 110  
 115  
 120  
 125  
 130  
 135  
 140  
 145  
 150  
 155  
 160  
 165  
 170  
 175  
 180  
 185  
 190  
 195

Met Val Asp Gln Leu Gln Gln Ile Leu Ser Val Ser Gln Leu Leu  
 Gln Lys His Gly Leu Gln Lys Pro Ile Ser Phe Val Lys Asn Thr  
 Gln Ser Ser Ser Ser Gln Ala Arg Lys Leu Met Val Arg Leu Thr  
 Arg Thr Leu Leu Gln Asp Met Leu Thr Met Gln Gln Asn Val Tyr  
 Thr Cys Leu Asp Ser Asp Ala Cys Tyr Gln Ile Phe Thr Gln Ser  
 Leu Leu Cys Ser Ser Arg Leu Gln Asn Ile His Leu Ala Gly Gln  
 Met Met His Cys Ser Ala Cys Ser Gln Asn Pro Pro Ala Gly Ile  
 Ala His Lys Gly Asn Pro His Tyr Arg Val Ser Tyr Gln Lys Ser  
 Ile Asp Leu Val Leu Ala Ser Arg Gln Tyr Phe Asn Ser Ser  
 Thr Asn Leu Thr Asp Ser Cys Met Asp Leu Ala Arg Cys Cys Leu  
 Gln Leu Ile Thr Asp Arg Pro Pro Ala Ile Gln Gln Leu Asp  
 Leu Ile Gln Ala Val Gly Cys Leu Gln Gln Phe Gly Val Lys Ile  
 Leu Pro Leu Gln Val Arg Leu Cys Pro Asp Arg Ile Ser Leu Ile





200	205	210
Lys Gln Cys Ile	Ser Gln Ser Pro Thr	Cys Tyr Lys Gln Ser
215	220	225
Lys Leu Leu Gly	Leu Ala Gln Leu Leu	Arg Val Ala Gly Gln
230	235	240
Pro Gln Gln Arg	Gly Gln Val Leu Ile	Leu Leu Val Gln Gln
245	250	255
Ala Leu Arg phe	His Asp Tyr Lys Ala	Ser Met His Cys Gln
260	265	270
Gln Leu Met Ala	Thr Gly Tyr Pro Lys	Ser Trp Asp Val Cys
275	280	285
Gln Leu Gly Gln	Ser Gln Gly Tyr Gln	Asp Leu Ala Thr Arg
290	295	300
Gln Leu Met Ala	Phe Ala Leu Thr His	Cys Pro Pro Ser Ser
305	310	315
Gln Leu Leu Leu	Ala Ser Ser Ser Ser	Leu Gln Thr Gln Ile
320	325	330
Tyr Gln Arg Val	Asn phe Gln Ile His	Gln Gly Gly Gln Asn
335	340	345
Ile Ser Ala Ser	Pro Leu Thr Ser Lys	Ala Val Gln Gln Asp
350	355	360
Val Gly Val Pro	Gly Ser Asn Ser Ala	Asp Leu Leu Arg Trp
365	370	375
Thr Ala Thr Thr	Met Lys Val Leu Ser	Asn Thr Thr Thr Thr
380	385	390
Lys Ala Val Leu	Gln Ala Val Ser Asp	Gly Gln Trp Trp Lys
395	400	405
Ser Leu Thr Tyr	Leu Arg Pro Leu Gln	Gly Gln Lys Cys Gly
410	415	420
Ala Tyr Gln Ile	Gly Thr Thr Ala Asn	Gln Asp Leu Gln Lys
425	430	435
Gly Cys His Pro	Phe Tyr Gln Ser Val	Ile Ser Asn Pro Phe
440	445	450
Ala Gln Ser Gln	Gly Thr Tyr Asp Thr	Gln His Val Pro Val
455	460	465
Gln Ser phe Ala	Gln Val Leu Leu Arg	Thr Gly Lys Leu Ala
470	475	480
Ala Lys Asn Lys	Gly Gln Val phe Pro	Thr Thr Gln Val Leu
485	490	495
Gln Leu Ala Ser	Ala Leu Pro Asn Asp	Met Thr Leu Ala Leu
500	505	510
Ala Tyr Leu Leu	Ala Leu Pro Gln Val	Leu Asp Ala Asn Arg
515	520	525
Phe Gln Lys Gln	Ser Pro Ser Ala Leu	Ser Leu Gln Leu Ala
530	535	540
Tyr Tyr Tyr Ser	Leu Gln Ile Tyr Ala	Arg Leu Ala Pro Cys
545	550	555
Arg Asp Lys Cys	His Pro Leu Tyr Arg	Ala Asp Pro Lys Gln
560	565	570
Ile Lys Met Val	Thr Arg His Val Thr	Arg His Gln His Gln
575	580	585
Trp Pro Gln Asp	Leu Ile Ser Leu Thr	Lys Gln Leu His Cys
590	595	600
Asn Gln Arg Leu	Leu Asp phe Thr Gln	Ala Gln Ile Leu Gln
605	610	615
Leu Arg Lys Gly	Val Asp Val Gln Arg	Phe Thr Ala Asp Gln
620	625	630
Tyr Lys Arg Gln	Thr Ile Leu Gly Leu	Ala Gln Thr Leu Gln
635	640	645
Ser Val Tyr Ser	Ile Ala Ile Ser Leu	Ala Gln Arg Tyr Ser
650	655	660
Ser Arg Trp Gln	Val phe Met Thr His	Leu Gln phe Leu phe
665	670	675



Asp Ser Gly Leu	Ser	Thr	Leu	Gln	Ile	Glu	Asn	Arg	Ala	Gln	Asp	680
Leu His Leu	Thr	Leu	Lys	Thr	Asp	Pro	Glu	Ala	Phe	His	Leu	690
Gln His Met	Val	Lys	Thr	Ile	Tyr	Pro	Tyr	Ile	Gly	Gly	Phe	705
His Gln Arg	Leu	Gln	Tyr	Phe	Thr	Leu	Gln	Asn	Cys	Gly	His	720
Cys Ala Asp	Leu	Gly	Asn	Cys	Ala	Ile	Lys	Pro	Gln	Thr	His	735
Arg Leu Leu	Lys	Phe	Lys	Val	Val	Ala	Ser	Gly	Leu	Asn	Tyr	750
Lys Leu Thr	Asp	Gln	Asn	Met	Ser	Pro	Leu	Gln	Ala	Leu	Gln	765
Pro Val Leu	Ser	Gln	Asn	Ile	Leu	Ser	Ile	Ser	Lys	Leu	Val	780
Pro Lys Ile	Pro	Gln	Lys	Asp	Gly	Gln	Met	Leu	Ser	Pro	Ser	795
Leu Tyr Thr	Ile	Trp	Leu	Gln	Lys	Phe	Trp	Thr	Gly	Asp	Pro	810
His Leu Ile	Lys	Gln	Val	Pro	Gly	Ser	Ser	Pro	Gln	Trp	Leu	825
Ala Tyr Asp	Val	Cys	Met	Lys	Tyr	Phe	Asp	Arg	Leu	His	Pro	840
Asp Leu Ile	Thr	Val	Val	Asp	Ala	Val	Thr	Ser	Pro	Lys	Ala	855
Val Thr Lys	Leu	Ser	Val	Glu	Ala	Arg	Lys	Glu	Met	Thr	Arg	870
Ala Ile Lys	Thr	Lys	Val	Lys	Phe	Ile	Glu	Lys	Pro	Arg	Lys	885
Asn Ser Glu	Asp	Glu	Ala	Lys	Asp	Ser	Lys	Val	Thr	Thr	Lys	900
Tyr Ala Asp	Thr	Leu	Asn	His	Leu	Glu	Lys	Ser	Leu	Ala	His	915
Glu Thr Leu	Ser	His	Ser	Phe	Ile	Leu	Ser	Lys	Asn	Ser	Glu	930
Gln Gln Thr	Leu	Gln	Lys	Tyr	Ser	His	Leu	Tyr	Asp	Leu	Ser	945
Ser Glu Lys	Glu	Lys	Leu	His	Asp	Glu	Ala	Val	Ala	Ile	Cys	960
Val Gly Pro	Leu	Asp	Ile	Ser	Pro	Lys	Asp	Ile	Val	Gln	Ser	990
Ile Met Lys	Ile	Ile	Ser	Ala	Leu	Ser	Gly	Gly	Ser	Ala	Asp	1005
Gly Gly Pro	Arg	Asp	Pro	Leu	Lys	Val	Leu	Gly	Val	Val	Ala	1020
Ala Val His	Ala	Ser	Val	Asp	Lys	Gly	Glu	Glu	Leu	Val	Ser	1035
Glu Asp Leu	Leu	Glu	Trp	Leu	Arg	Pro	Phe	Cys	Ala	Asp	Ala	1050
Trp pro Val	Arg	Pro	Arg	Ile	His	Val	Leu	Gln	Ile	Leu	Gly	1065
Ser Phe His	Leu	Thr	Glu	Glu	Asp	Ser	Lys	Leu	Leu	Val	Phe	1080
Arg Thr Glu	Ala	Ile	Leu	Lys	Ala	Ser	Trp	Pro	Gln	Arg	Gln	1095
Asp Ile Ala	Asp	Ile	Glu	Asn	Glu	Glu	Asn	Arg	Tyr	Cys	Leu	1110
Met Glu Leu	Leu	Glu	Ser	His	His	Glu	Ala	Glu	Phe	Gln	His	1125
Leu Val Leu	Leu	Gln	Ala	Trp	Pro	Pro	Met	Lys	Ser	Glu	Tyr	1140



1145 Val Ile Thr Asn Asn Pro Trp Val Arg Leu Ala Thr Val Met Leu 1150  
 1160 Thr Arg Cys Thr Met Glu Asn Lys Glu Gly Leu Gly Asn Glu Val 1165  
 1175 Thr Arg Cys Thr Met Glu Asn Lys Glu Gly Leu Gly Asn Glu Val 1180  
 1185 Leu Lys Met Cys Arg Ser Leu Tyr Asn Thr Lys Glu Met Leu Pro 1195  
 1190 Ala Glu Gly Val Lys Glu Leu Cys Leu Leu Leu Asn Glu Ser 1200  
 1205 Ala Glu Gly Val Lys Glu Leu Cys Leu Leu Leu Asn Glu Ser 1210  
 1215 Leu Leu Leu Pro Ser Leu Lys Leu Leu Leu Glu Ser Arg Asp Glu 1220  
 1220 His Leu His Glu Met Ala Leu Glu Ile Thr Ala Val Thr Thr 1225  
 1235 Val Asn Asp Ser Asn Cys Asp Glu Leu Ser Leu Leu Leu 1240  
 1245 Val Asn Asp Ser Asn Cys Asp Glu Leu Ser Leu Leu Leu 1250  
 1255 Asp Ala Lys Leu Leu Val Lys Cys Val Ser Thr Pro Phe Tyr Pro 1260  
 1265 Arg Ile Val Asp His Leu Leu Ala Ser Leu Glu Ile Glu Gly Arg Trp 1270  
 1275 Arg Ile Val Asp His Leu Leu Ala Ser Leu Glu Ile Glu Gly Arg Trp 1285  
 1280 Asp Ala Glu Glu Leu Gly Arg His Leu Arg Glu Ala Gly His Glu 1290  
 1295 Ala Glu Ala Gly Ser Leu Leu Leu Ala Val Arg Gly Thr His Glu 1300  
 1310 Ala Phe Arg Thr Phe Ser Thr Ala Leu Arg Ala Ala Glu His Trp 1315  
 1325 Val 1330

<210> 38  
 <211> 934  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2259032CD1  
 <400> 38

Met Phe Thr Trp Lys Phe Asp Leu Asn Thr Thr Ser His Val Asp Lys 1  
 5 Leu Leu Asp Lys Glu His Val Thr Leu Glu 10  
 20 Asp Asp Ile Leu Glu Glu Cys Lys Ala Glu Asn Glu Lys Leu Leu 25  
 35 Asp Phe Leu Cys Arg Glu Glu Cys Met Glu Glu Leu Val Ser Leu 40  
 50 Ile Thr Glu Asp Pro Leu Asp Met Glu Lys Val Arg Phe 45  
 60 Lys Tyr Pro Asn Thr Ala Cys Glu Leu Leu Thr Cys Asp Val Pro 55  
 75 Ala Glu Asp Thr Ala Cys Glu Leu Leu Thr Cys Asp Val Pro 60  
 80 Glu Ile Ser Asp Arg Leu Gly Gly Asp Glu Ser Leu Leu Ser Leu 65  
 95 Leu Tyr Asp Phe Leu Asp His Glu Pro Leu Asn Pro Leu Leu 70  
 110 Ala Ser Phe Phe Ser Lys Thr Ile Gly Asn Leu Ile Ala Arg Lys 75  
 125 Thr Glu Glu Val Ile Thr Phe Leu Lys Lys Asp Lys Phe Ile 80  
 140 Ser Leu Val Leu Lys His Ile Gly Thr Ser Ala Leu Met Asp Leu 85  
 155 Leu Leu Arg Leu Val Ser Cys Val Glu Pro Ala Gly Leu Arg Glu 90  
 170 Asp Val Leu His Trp Leu Asn Glu Glu Lys Val Ile Glu Arg Leu 95  
 185



Ala	Pro	Gly	Ala	Gly	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

.

.

.

.



665 Pro pro Val Glu Gly Asp Ser Glu Ala Gly Ala Met Trp Thr Ala 670  
680 Val phe Asp Glu Pro Ala Asn Ser Thr 700  
695 Val Arg Asp Val Gly Ser Ser Val Trp Ala 715  
710 Pro Glu Glu Lys Gly Trp Ala Lys phe Thr 730  
725 Cys Cys Ser Ser Glu Gly Pro Arg Cys Ser 745  
740 Glu Cys Ser Ser His Ala Glu Gly Ser Arg Ser 760  
755 Ala phe Ser Ser Pro Ala Ser Pro Cys Ala Trp 775  
780 Arg Lys Ala Pro Leu Ala Ser Ser Asp Ser 790  
795 Ser His Ser Glu Asp Gly Asp Gln Lys Ala 805  
800 Ala Val Ser Arg Gly Pro Gly Arg Glu Ala 820  
815 Val Ala Arg Thr Glu Ala Val Gly Arg Val Gly Cys Ala Asp 840  
830 Ser Arg Leu Leu Ser Pro Ala Cys Pro Ala 850  
845 Ala Ala Pro Ala Val Ala Val pro pro Glu Ala Thr Val Ala Ile 865  
860 Thr Thr Ala Leu Ser Lys Ala Gly Pro Ala Ile pro Thr Pro Ala 885  
885 Val Ser Ser Ala Leu Ala Val Ala Val pro Leu Gly Pro Ile Met 900  
890 Ala Val Thr Ala Ala Pro Ala Met Val Ala Thr Leu Gly Thr Val 915  
905 Thr Lys Asp Gly Lys Thr Asp Ala pro pro Glu Gly Ala Ala Leu 930  
Asn Gly pro Val

<210> 39

<211> 515

<212> PRT

<213> Homo sapiens

<220>

<221> misc-feature

<223> Incyte ID No: 2359526CD1

<400> 39

Met Ala Ala Asn Met Tyr Arg Val Gly Asp Tyr Val Tyr phe Glu 1  
1 Asn Ser Ser Ser Asn Pro Tyr Leu Ile Arg Arg Ile Glu Glu Leu 5  
20 Asn Lys Thr Ala Ser Gly Asn Val Glu Ala Lys Val Val Cys phe 25  
35 Tyr Arg Arg Arg Asp Ile Ser Asn Thr Leu Ile Met Leu Ala Asp 40  
50 Lys His Ala Lys Glu Ile Glu Glu Ser Glu Thr Thr Val Glu 55  
65 Ala Asp Leu Thr Asp Lys Gln Lys His Gln Lys His Arg Glu 70  
80 Leu phe Leu Ser Arg Gln Tyr Glu Ser Leu Pro Ala Thr His Ile 85  
95 Arg Gly Lys Cys Ser Val Ala Leu Leu Asn Glu Thr Glu Ser Val 100  
110 Arg Gly Lys Cys Ser Val Ala Leu Leu Asn Glu Thr Glu Ser Val 115  
120



37/93

Leu Ser Tyr Leu Asp Lys Gln Asp Thr Phe Tyr Ser Leu Val  
125  
130  
135  
Tyr Asp Pro Ser Leu Lys Thr Leu Leu Ala Asp Lys Gln Ile  
140  
145  
150  
Arg Val Gly Pro Arg Gln Ala Asp Ile Pro Gln Met Leu  
155  
160  
165  
Gln Gly Gln Ser Asp Gln Arg Gln Lys Leu Gln Val Lys  
170  
175  
180  
Val Trp Asp Pro Asn Ser Pro Leu Thr Asp Arg Gln Ile Asp  
185  
190  
195  
Phe Leu Val Val Ala Arg Ala Val Gly Thr Phe Ala Arg Ala  
200  
205  
210  
Asp Cys Ser Ser Ser Val Arg Gln Pro Ser Leu His Met Ser  
215  
220  
225  
Ala Ala Ala Ser Arg Asp Ile Thr Leu Phe His Ala Met Asp  
230  
235  
240  
Leu Tyr Arg His Tyr Asp Leu Ser Ser Ala Ile Ser Val Leu  
245  
250  
255  
Val Pro Leu Gly Gly Pro Val Leu Cys Arg Gln Met Gln Gln  
260  
265  
270  
Trp Ser Ala Ser Gln Ala Ser Leu Phe Gln Ala Leu Gln Lys  
275  
280  
285  
Tyr Gly Lys Asp Phe Asn Asp Ile Arg Gln Asp Phe Leu Pro  
290  
295  
300  
Lys Ser Leu Thr Ser Ile Ile Gln Tyr Tyr Tyr Met Trp Lys Thr  
305  
310  
315  
Thr Asp Arg Tyr Val Gln Gln Lys Arg Leu Lys Ala Ala Gln  
320  
325  
330  
Gln Ser Lys Leu Lys Gln Val Tyr Ile Pro Thr Tyr Ser Lys Pro  
335  
340  
345  
Asn Pro Asn Gln Ile Ser Thr Ser Asn Gly Lys Pro Gly Ala Val  
350  
355  
360  
Asn Gly Ala Val Gly Thr Thr Phe Gln Pro Gln Asn Pro Leu  
365  
370  
375  
Gly Arg Ala Cys Gln Ser Cys Tyr Ala Thr Gln Ser His Gln Trp  
380  
385  
390  
Tyr Ser Trp Gly Pro Asn Met Gln Cys Arg Leu Cys Ala Ile  
395  
400  
405  
Cys Trp Leu Tyr Trp Lys Lys Tyr Gly Gly Leu Lys Met Pro Thr  
410  
415  
420  
Gln Ser Gln Gln Lys Leu Ser Pro Ser Pro Thr Thr Gln Asp  
425  
430  
435  
Pro Arg Val Arg Ser His Val Ser Arg Gln Ala Met Gln Gly Met  
440  
445  
450  
Pro Val Arg Asn Thr Gly Ser Pro Lys Ser Ala Val Lys Thr Arg  
455  
460  
465  
Gln Ala Phe Phe His Thr Thr Tyr Phe Thr Lys Phe Ala Arg  
470  
475  
480  
Gln Val Cys Lys Asn Thr Leu Arg Leu Arg Gln Ala Arg Arg  
485  
490  
495  
Pro Phe Val Ala Ile Asn Tyr Ala Ala Ile Arg Ala Gln Cys Lys  
500  
505  
510  
Met Leu Leu Asn Ser

<210> 40  
<211> 146  
<212> PRT  
<213> Homo sapiens  
<220>  
<221> misc\_feature  
<223> Incyte ID No: 2456494CD1



<400> 40  
Met Val Asp Glu Leu Val Leu Leu Leu His Ala Leu Leu Met Arg  
1  
His Arg Ala Leu Ser Ile Glu Asn Ser Glu Met Glu Glu Leu  
20  
Arg Leu Leu Val Cys Glu Arg Ala Ser Leu Leu Arg Glu Val Arg  
35  
Pro Pro Ser Cys Pro Val Pro Phe Pro Glu Thr Phe Asn Gly Glu  
50  
Ser Ser Arg Leu Pro Glu Phe Ile Val Glu Thr Ala Ser Tyr Met  
65  
Leu Val Asn Glu Asn Arg Phe Cys Asn Asp Ala Met Lys Val Ala  
80  
Phe Leu Ile Ser Leu Leu Thr Gly Glu Ala Glu Glu Trp Val Val  
95  
Pro Tyr Ile Glu Met Asp Ser Pro Ile Leu Glu Asp Tyr Arg Ala  
110  
Phe Leu Asp Glu Met Lys Glu Cys Phe Gly Trp Asp Asp Glu  
125  
Asp Asp Asp Asp Glu Glu Glu Glu Glu Asp Tyr  
140  
145

<210> 41  
<211> 580  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc-feature  
<223> Incyte ID No: 2668536CD1

<400> 41  
Met Lys Glu Asn Lys Glu Asn Ser Ser Pro Ser Val Thr Ser Ala  
1  
Met Leu Asp His Thr Lys Pro Cys Trp Tyr Trp Asp Lys Asp  
20  
Leu Ala His Thr Pro Ser Glu Leu Glu Gly Leu Asp Pro Ala Thr  
35  
Glu Ala Arg Tyr Arg Glu Gly Ala Arg Phe Ile Phe Asp Val  
50  
Gly Thr Arg Leu Glu His Tyr Asp Thr Leu Ala Thr Gly Ile  
65  
Ile Tyr Phe His Arg Phe Tyr Met Phe His Ser Phe Lys Glu Phe  
80  
Pro Arg Tyr Val Thr Gly Ala Cys Cys Leu Phe Leu Ala Gly Lys  
95  
Val Glu Glu Thr Pro Lys Lys Cys Lys Asp Ile Ile Lys Thr Ala  
110  
Arg Ser Leu Leu Asn Asp Val Glu Phe Gly Glu Phe Gly Asp  
125  
Pro Lys Glu Glu Val Met Val Leu Glu Arg Ile Leu Leu Glu Thr  
140  
Ile Lys Phe Asp Leu Glu Val Glu His Tyr Glu Phe Leu Leu  
155  
Lys Tyr Ala Lys Glu Leu Lys Gly Asp Lys Asn Lys Ile Glu Lys  
170  
Leu Val Glu Met Ala Trp Thr Phe Val Asn Asp Ser Leu Cys Thr  
185  
Thr Leu Ser Leu Glu Trp Glu Pro Glu Ile Ala Val Ala Val  
200  
Met Tyr Leu Ala Gly Arg Leu Cys Lys Phe Glu Ile Glu Glu Trp  
215  
Thr Ser Lys Pro Met Tyr Arg Arg Trp Trp Glu Glu Phe Val Glu  
230  
235



Asp Val Pro Val Asp Val Leu Gln Asp Ile Cys His Gln Ile Leu 255  
 Asp Leu Tyr Ser Gln Gly Lys Gln Gln Met Pro His His Thr Pro 270  
 His Gln Leu Gln Gln Pro Pro Ser Leu Gln Pro Thr Pro Gln Val 285  
 Pro Gln Val Gln Gln Ser Gln Pro Ser Gln Ser Ser Gln Pro Ser 300  
 Gln Pro Gln Gln Lys Asp Pro Gln Gln Pro Ala Gln Gln Gln Gln 315  
 Pro Ala Gln Gln Pro Lys Lys Lys Pro Ser Pro Gln Pro Ser Ser Pro 330  
 Arg Gln Val Lys Arg Ala Val Val Val Ser Pro Lys Gln Gln Asp 345  
 Lys Ala Ala Gln Pro Pro Pro Pro Lys Ile Pro Lys Ile Gln Thr 360  
 Thr His Pro Pro Leu Pro Pro Ala His Pro Pro Pro Asp Arg Lys 375  
 Pro Pro Leu Ala Ala Leu Gln Gly Gln Ala Gln Pro Pro Gly Pro 390  
 Val Asp Ala Thr Asp Leu Pro Lys Val Gln Ile Pro Pro Pro Ala 405  
 His Pro Ala Pro Val His Gln Pro Pro Pro Leu Pro His Arg Pro 420  
 Pro Pro Pro Pro Ser Ser Tyr Met Thr Gly Met Ser Thr Thr 435  
 Ser Ser Tyr Met Ser Gly Gln Gly Tyr Gln Ser Leu Gln Ser Met 450  
 Met Lys Thr Gln Gly Pro Ser Tyr Gly Ala Leu Pro Pro Ala Tyr 465  
 Gly Pro Pro Ala His Leu Pro Tyr His Pro His Val Tyr Pro Pro 480  
 Asn Pro Pro Pro Pro Val Pro Pro Pro Pro Ala Ser Phe Pro 495  
 His Leu Pro Ser His Pro Leu Leu Leu Ala Thr Pro Asn Pro His 510  
 Pro Pro Thr Thr Pro Thr Ser His Pro His Pro His Ala Ser Arg 525  
 Leu Pro Thr Gln Ser Pro Leu Ile Leu Leu Gln Gly Trp Ala Cys 540  
 Arg Gln Pro Ala Thr His Leu Leu Pro Ser Pro Leu Gln Asp Ser 555  
 Leu Leu Cys Pro Arg Pro Phe Pro His Pro Ala Cys Leu Gln Leu 570  
 Gln Gly Leu Gly Arg Ala Ala Trp Met Arg 580  
 575  
 560  
 545  
 530  
 515  
 500  
 485  
 470  
 455  
 440  
 425  
 410  
 395  
 380  
 365  
 350  
 335  
 320  
 305  
 290  
 275  
 260  
 245  
 230  
 215  
 200  
 185  
 170  
 155  
 140  
 125  
 110  
 95  
 80  
 65  
 50  
 35  
 20  
 5  
 1  
 Met Ala Gln Pro Asp Tyr Ile Gln Asp Asp Asn Pro Gln Leu Ile  
 Arg Pro Gln Lys Leu Ile Asn Pro Val Lys Thr Ser Arg Asn His  
 Gln Asp Leu His Arg Gln Leu Leu Met Asn Gln Lys Arg Gly Leu  
 Ala Pro Gln Asn Lys Pro Gln Leu Gln Lys Val Met Gln Lys Arg

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 2683225CD1

<210> 42  
 <211> 131  
 <212> PRT  
 <213> Homo sapiens





```
<210> 43  
<211> 812  
<212> PRT  
<213> Homo sapiens
```

<400> 43  
Met Gly Arg Lys Leu Asp Pro Thr Lys Glu Lys Arg Gly Pro Gly  
1  
5  
10  
15

40/93



Thr	Leu	Lys	Thr	Arg	Arg	Arg	Leu	Ala	Gln	Ala	Leu	Ile	Asn	330
Arg	Gly	Val	Asn	Asp	Pro	Leu	Gly	Lys	Trp	Ser	Lys	Thr	Gly	345
Leu	Val	Val	Tyr	Asp	Ser	Ser	Val	Pro	Ile	Gly	Ala	Thr	Pro	360
Tyr	Leu	Ala	Gly	His	Tyr	Met	Leu	Gln	Gly	Ala	Ser	Ser	Met	375
Pro	Val	Met	Ala	Leu	Ala	Pro	Gln	Gln	His	Glu	Arg	Ile	Leu	390
Met	Cys	Cys	Ala	Pro	Gly	Gly	Lys	Thr	Ser	Tyr	Met	Ala	Gln	405
Met	Lys	Asn	Thr	Gly	Val	Ile	Leu	Ala	Asn	Asp	Ala	Asn	Ala	420
Arg	Leu	Lys	Ser	Val	Gly	Asn	Leu	His	Arg	Leu	Gly	Val	Thr	435
Asn	Thr	Ile	Ile	Ser	His	Tyr	Asp	Gly	Arg	Gln	Phe	Pro	Lys	450
Val	Gly	Gly	Phe	Asp	Arg	Val	Leu	Leu	Asp	Ala	Pro	Cys	Ser	465
Thr	Gly	Val	Ile	Ser	Lys	Asp	Pro	Ala	Val	Lys	Thr	Asn	Lys	480
Glu	Lys	Asp	Ile	Leu	Arg	Cys	Ala	His	Leu	Gln	Lys	Glu	Leu	495
Leu	Ser	Ala	Ile	Asp	Ser	Val	Asn	Ala	Thr	Ser	Lys	Thr	Gly	510
Tyr	Leu	Val	Tyr	Cys	Thr	Cys	Ser	Ile	Thr	Val	Glu	Glu	Asn	525
Trp	Val	Val	Asp	Tyr	Ala	Leu	Lys	Lys	Arg	Asn	Val	Arg	Leu	540
Pro	Thr	Gly	Leu	Asp	Phe	Gly	Gln	Glu	Gly	Phe	Thr	Arg	Phe	555
Glu	Arg	Arg	Phe	His	Pro	Ser	Leu	Arg	Ser	Thr	Arg	Arg	Phe	570
Pro	His	Thr	His	Asn	Met	Asp	Gly	Phe	Phe	Ile	Ala	Lys	Phe	585
Lys	Phe	Ser	Asn	Ser	Ile	Pro	Gln	Ser	Gln	Thr	Gly	Asn	Ser	600
Thr	Ala	Thr	Pro	Thr	Asn	Val	Asp	Leu	Pro	Gln	Val	Ile	Pro	615
Ser	Glu	Asn	Ser	Ser	Gln	Pro	Ala	Lys	Lys	Ala	Lys	Gly	Ala	630
Lys	Thr	Lys	Gln	Gln	Leu	Gln	Lys	Gln	Gln	His	Pro	Lys	Lys	645
Ser	Phe	Gln	Lys	Leu	Asn	Gly	Ile	Ser	Lys	Gly	Ala	Asp	Ser	660
Leu	Ser	Thr	Val	Pro	Ser	Val	Thr	Lys	Thr	Gln	Ala	Ser	Ser	675
Phe	Gln	Asp	Ser	Ser	Gln	Pro	Ala	Gly	Lys	Ala	Glu	Gly	Ile	690
Glu	Pro	Lys	Val	Thr	Gly	Lys	Leu	Lys	Gln	Arg	Ser	Pro	Lys	705
Gln	Ser	Ser	Lys	Lys	Val	Ala	Phe	Leu	Arg	Gln	Asn	Ala	Pro	720
Lys	Gly	Thr	Asp	Thr	Gln	Thr	Pro	Ala	Val	Leu	Ser	Pro	Ser	735
Thr	Gln	Ala	Thr	Leu	Lys	Pro	Lys	Asp	His	His	Gln	Pro	Leu	750
Arg	Ala	Lys	Gly	Val	Glu	Lys	Gln	Gln	Leu	Pro	Glu	Gln	Pro	765
Glu	Lys	Ala	Ala	Phe	Gln	Lys	Gln	Asn	Asp	Thr	Pro	Lys	Gly	780
Gln	Pro	Pro	Thr	Val	Ser	Pro	Ile	Arg	Ser	Ser	Arg	Pro	Pro	



785 Ala Lys Arg Lys Ser Gln Ser Arg Gly Asn Ser Gln Leu 795  
 800  
 Leu Ser 810

<210> 44

<211> 537

<212> PRT

<213> Homo sapiens

<220>

<221> misc-feature

<223> Incyte ID No: 2959521CD1

<400> 44

Met Arg Gly Val Gly Ala Arg Val Tyr Ala Asp Ala Pro Ala Lys  
 1  
 Leu Leu Pro Pro Pro Ala Ala Trp Asp Leu Ala Val Arg Leu  
 20  
 35 Arg Gly Ala Gln Ala Ser Gln Arg Gln Val Tyr Ser Val Thr  
 40  
 50 Met Lys Leu Leu Leu Leu His Pro Ala Phe Gln Ser Cys Leu Leu  
 55  
 60 Leu Thr Leu Leu Gly Leu Trp Arg Thr Thr Pro Gln Ala His Ala  
 70  
 65 Ser Ser Leu Gly Ala Pro Ala Ile Ser Ala Ala Ser Phe Leu Gln  
 80  
 85 Asp Leu Ile His Arg Tyr Gly Gly Asp Ser Leu Thr Leu Gln  
 90  
 95 Arg Leu Ile His Arg Tyr Gly Gly Asp Ser Leu Thr Leu Gln  
 100  
 110 Gln Leu Lys Ala Leu Leu Asn His Leu Asp Val Gly Val Gly Arg  
 115  
 125 Gly Asn Val Thr Gln His Val Gln Gly His Arg Asn Leu Ser Thr  
 130  
 140 Cys Phe Ser Ser Gly Asp Leu Phe Thr Ala His Asn Phe Ser Gln  
 145  
 150 Gln Ser Arg Ile Gly Ser Ser Gln Leu Gln Phe Cys Pro Thr  
 155  
 160 Ile Leu Gln Gln Leu Asp Ser Arg Ala Cys Thr Ser Gln Asn Gln  
 165  
 170 Gln Asn Gln Leu Leu Leu Tyr Phe Ile Ala Leu Ala Ile  
 175  
 185 Gln Asn Gln Gln Thr Gln Thr Gln Gly Arg Pro Ser Ala  
 190  
 200 Val Gln Val Trp Gly Tyr Gly Leu Leu Cys Val Thr Val Ile Ser  
 205  
 215 Leu Cys Ser Leu Leu Gly Ala Ser Val Val Pro Phe Met Lys Lys  
 220  
 230 Thr Phe Tyr Lys Arg Leu Leu Leu Tyr Phe Ile Ala Leu Ala Ile  
 235  
 245 Gly Thr Leu Tyr Ser Asn Ala Leu Phe Gln Leu Ile Pro Gln Ala  
 250  
 255 Phe Gly Phe Asn Pro Leu Gln Asp Tyr Tyr Val Ser Lys Ser Ala  
 260  
 275 Val Val Phe Gly Phe Tyr Leu Phe Phe Thr Gln Lys Ile  
 280  
 285 Leu Lys Ile Leu Leu Lys Gln Lys Asn Gln His His His Gly His  
 295  
 305 Ser His Tyr Ala Ser Gln Ser Leu Pro Ser Lys Lys Asp Gln Gln  
 310  
 320 Gln Gly Val Met Gln Lys Leu Gln Asn Gly Asp Leu Asp His Met  
 325  
 335 Ile Pro Gln His Cys Ser Ser Gln Leu Asp Gly Lys Ala Pro Met  
 340  
 350 Val Asp Gln Lys Val Ile Val Gly Ser Leu Ser Val Gln Asp Leu  
 355  
 360



Gln Ala Ser Gln	Ser	Ala	Cys	Tyr	Trp	Leu	Lys	Gly	Val	Arg	Tyr	365
Ser Asp Ile Gly	Thr	Leu	Ala	Trp	Met	Ile	Thr	Leu	Ser	Asp	Gly	375
Leu His Asn Phe	Ile	Asp	Gly	Leu	Ala	Ile	Gly	Ala	Ser	Phe	Thr	390
Val Ser Val Phe	Gln	Gly	Ile	Ser	Thr	Ser	Val	Ala	Ile	Leu	Cys	405
Ala Gly Met Ser	Ile	Gln	Gln	Ala	Leu	Phe	Phe	Asn	Phe	Leu	Ser	420
Glu Gln Phe Pro	His	Gln	Leu	Gly	Asp	Phe	Val	Ile	Leu	Leu	Asn	435
Ala Cys Cys Cys	Tyr	Leu	Gly	Leu	Ala	Phe	Gly	Ile	Leu	Ala	Gly	450
Ser His Phe Ser	Ala	Asn	Trp	Ile	Phe	Ala	Leu	Ala	Gly	Gly	Met	465
Phe Leu Tyr Ile	Ser	Leu	Ala	Asp	Met	Phe	Pro	Gln	Met	Asn	Gln	480
Val Cys Gln Gln	Asp	Gln	Arg	Lys	Gly	Ser	Ile	Leu	Ile	Pro	Phe	495
Ile Ile Gln Asn	Leu	Gly	Leu	Leu	Thr	Gly	Phe	Thr	Ile	Met	Val	510
Val Leu Thr Met	Tyr	Ser	Gly	Gln	Ile	Gln	Ile	Gly				525
												535

<210> 45  
 <211> 584  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 3082014CD1

Met Leu Trp Gly	Gly	Arg	Val	Gly	Leu	Thr	Gly	Val	Phe	Gln	Ser	1
Leu Ser Tyr Arg	Gly	Lys	Cys	Ser	Val	Thr	Leu	Leu	Asn	Gln	Thr	15
Asp Ile Leu Ser	Gln	Tyr	Leu	Gln	Lys	Gln	Lys	Asp	Cys	Phe	Phe	30
Ser Leu Val Phe	Asp	Pro	Val	Gln	Lys	Thr	Leu	Leu	Ala	Asp	Gln	45
Ser Leu Val Phe	Asp	Pro	Val	Gln	Lys	Thr	Leu	Leu	Ala	Asp	Gln	60
Gly Gln Ile Arg	Val	Gly	Cys	Lys	Tyr	Gln	Ala	Gln	Ile	Pro	Asp	75
Arg Leu Val Gln	Gly	Gln	Ser	Asp	Asn	Arg	Asn	Gln	Gln	Lys	Met	90
Glu Met Lys Val	Trp	Asp	Pro	Asp	Asn	Pro	Leu	Thr	Asp	Arg	Gln	105
Ile Asp Gln Phe	Leu	Val	Val	Ala	Arg	Ala	Val	Gly	Thr	Phe	Ala	120
Arg Ala Leu Asp	Cys	Ser	Ser	Ser	Ile	Arg	Gln	Pro	Ser	Leu	His	135
Met Ser Ala Ala	Ala	Ser	Arg	Asp	Ile	Thr	Leu	Phe	His	Ala		150
Met Asp Thr Leu	Gln	Arg	Asn	Gly	Tyr	Asp	Leu	Ala	Lys	Ala	Met	165
Ser Thr Leu Val	Pro	Gln	Gly	Gly	Pro	Val	Leu	Cys	Arg	Asp	Gln	180
Met Gln Gln Trp	Ser	Ala	Ser	Gln	Ala	Met	Leu	Phe	Gln	Ala		195
Leu Gln Lys Tyr	Gly	Lys	Asp	Phe	Asn	Asp	Ile	Arg	Gln	Asp	Phe	210
Leu Pro Trp Lys	Ser	Leu	Ala	Ser	Ile	Val	Gln	Phe	Tyr	Tyr	Met	220





225  
 215 Arg Tyr Ile Gln Gln Lys Arg Lys Ala  
 230 Trp Lys Thr Thr Asp  
 235 Ala Gln Ala Asp Ser Lys Leu Lys Gln Val Tyr Ile Pro Thr  
 240  
 245 Thr Lys Pro Asn Pro Asn Gln Ile Ile Ser Val Gly Ser Lys Pro  
 250  
 255 Gly Met Asn Gly Ala Gly phe Gln Lys Gly Leu Thr Cys Gln Ser  
 260  
 265 Cys His Thr Thr Gln Ser Ala Gln Trp Tyr Ala Trp Gly Pro  
 270  
 275 Lys Met Gln Cys Arg Leu Cys Ala Ser Cys Trp Ile Tyr Trp Lys  
 280  
 285 Lys Tyr Gly Gly Leu Lys Thr Pro Thr Gln Leu Gly Ala Thr  
 290  
 295 Lys Leu Thr Arg Leu Ala Arg Arg Met Cys Arg Asp Leu Leu Gln  
 300  
 305 Pro Arg Arg Ala Ala Arg Pro Tyr Ala Pro Ile Asn Ala Asn  
 310  
 315 Ala Ile Lys Ala Gln Cys Ser Ile Arg Leu Pro Lys Ala Ala Lys  
 320  
 325 Thr Pro Leu Lys Ile His Pro Leu Val Arg Leu Pro Leu Ala Thr  
 330  
 335 Ile Val Lys Asp Leu Val Ala Gln Ala Pro Leu Lys Pro Lys Thr  
 340  
 345 Pro Arg Gly Thr Thr Gln Thr Lys Thr Pro Ile Asn Ala Asn  
 350  
 355 Lys Leu Thr Arg Leu Ala Arg Arg Met Cys Arg Asp Leu Leu Gln  
 360  
 365 Thr Arg Gly Thr Lys Thr Pro Ile Asn Arg Asn Gln Leu Ser Gln  
 370  
 375 Asn Arg Gly Leu Met Val Lys Arg Ala Tyr Gln Thr  
 380  
 385 Met Ala Gly Ala Gly Val Pro Phe Ser Ala Asn Gly Arg Pro Leu  
 390  
 395 Ser His Pro Ala Ser Ser Ser Gln Pro Ala Ala Lys Arg Gln  
 400  
 405 Thr Lys Asp Thr Lys Thr Pro Asn Pro Val Val Phe Val Ala  
 410  
 415 Thr Lys Asp Thr Arg Ala Leu Arg Lys Ala Leu Thr His Leu Gln  
 420  
 425 Met Arg Arg Ala Ala Arg Arg Pro Asn Leu Pro Lys Val Lys  
 430  
 435 Pro Thr Leu Ile Arg Ser Ser Ser Gln Pro Ala Ala Lys Arg Gln  
 440  
 445 Thr Lys Asp Thr Arg Ala Leu Arg Lys Ala Leu Thr His Leu Gln  
 450  
 455 Met Arg Arg Ala Ala Arg Arg Pro Asn Leu Pro Lys Val Lys  
 460  
 465 Pro Thr Leu Ile Arg Ser Ser Ser Ser Gln Pro Ala Ala Lys Arg Gln  
 470  
 475 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 480  
 485 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 490  
 495 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 500  
 505 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 510  
 515 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 520  
 525 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 530  
 535 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 540  
 545 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 550  
 555 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 560  
 565 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 570  
 575 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 580  
 585 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 590  
 595 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 600  
 605 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 610  
 615 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 620  
 625 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 630  
 635 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 640  
 645 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 650  
 655 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 660  
 665 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 670  
 675 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 680  
 685 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 690  
 695 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 700  
 705 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 710  
 715 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 720  
 725 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 730  
 735 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 740  
 745 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 750  
 755 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 760  
 765 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 770  
 775 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 780  
 785 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 790  
 795 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 800  
 805 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 810  
 815 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 820  
 825 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 830  
 835 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 840  
 845 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 850  
 855 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 860  
 865 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 870  
 875 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 880  
 885 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 890  
 895 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 900  
 905 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 910  
 915 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 920  
 925 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 930  
 935 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 940  
 945 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 950  
 955 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 960  
 965 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 970  
 975 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 980  
 985 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 990  
 995 Ser His Pro Ala Ser Thr Asn Gln Pro Ile Val Leu Gln Asp  
 1000

<210> 46  
 <211> 425  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc-feature  
 <223> Incyte ID No: 3520701CD1  
 <400> 46

Met Ala Gly Ala Gln Gly Ala Ala Gly Arg Gln Ser Gln Leu Gln  
 1  
 5  
 10  
 15  
 20  
 25  
 30



Gln Asn Gln Ala Cys Ala Val Leu Gly Gly Ser Asp Ser Gln Lys  
 35  
 Cys Ser Tyr Ser Gln Gly Ser Val Lys Arg Gln Ala Leu Tyr Ala  
 40  
 Cys Ser Tyr Ser Gln Gly Ser Val Lys Arg Gln Ala Leu Tyr Ala  
 45  
 Leu Tyr Thr Cys Thr Pro Gln Gly Gln Cys His Lys Leu Phe Gln  
 50  
 Cys Ser Thr Cys Thr Pro Gln Gly Gln Cys His Lys Leu Phe Gln  
 55  
 Phe Lys Asn Leu Gln Cys Lys Leu Leu Pro Asp Lys Ala Lys Val  
 60  
 Asn Ser Gly Asn Lys Tyr Asn Asp Asn Phe Gly Leu Tyr Cys  
 65  
 Ile Cys Lys Arg Pro Tyr Pro Asp Pro Gln Asp Gln Ile Pro Asp  
 70  
 Gln Met Ile Gln Cys Val Val Cys Gln Asp Trp Phe His Gly Arg  
 75  
 His Leu Gly Ala Ile Pro Pro Gln Ser Gly Asp Phe Gln Gln Met  
 80  
 Val Cys Gln Ala Cys Met Lys Arg Cys Ser Phe Leu Trp Ala Tyr  
 85  
 Ala Ala Gln Leu Ala Val Thr Lys Ile Ser Thr Gln Asp Asp Gly  
 90  
 Leu Val Arg Asn Ile Asp Gly Ile Gly Asp Gln Gln Val Ile Lys  
 95  
 Pro Gln Asn Gly Gln His Gln Asp Ser Thr Leu Lys Gln Asp Val  
 100  
 Pro Gln Gln Gly Lys Asp Asp Val Arg Gln Val Lys Val Gln Gln  
 105  
 Asn Ser Gln Pro Cys Ala Gly Ser Ser Ser Ser Ser Ser Ser Ser  
 110  
 Thr Val Phe Lys Asn Gln Ser Leu Asn Ala Gln Ser Lys Ser Gly  
 115  
 Cys Lys Leu Gln Gln Leu Lys Ala Lys Gln Ile Lys Lys Lys Asp  
 120  
 Thr Ala Thr Tyr Trp Pro Leu Asn Trp Arg Ser Lys Leu Cys Thr  
 125  
 Cys Gln Asp Cys Met Lys Met Tyr Gly Asp Leu Asp Val Leu Phe  
 130  
 Leu Thr Asp Gln Tyr Asp Thr Val Leu Ala Tyr 325  
 320  
 330  
 335  
 340  
 345  
 Lys Ile Ala Gln Ala Thr Asp Arg Ser Asp Pro Leu Met Asp Thr  
 350  
 Leu Ser Ser Ser Met Asn Arg Val Gln Gln Val Gln Leu Ile Cys Gln  
 355  
 Tyr Asn Asp Leu Lys Thr Gln Lys Arg Phe 370  
 375  
 Ala Asp Gln Gly Thr Val Val Lys Arg Gln Asp Ile Gln Gln Phe  
 380  
 Phe Gln Gln Phe Gln Ser Lys Lys Arg Arg Val Asp Gly Met  
 385  
 Gln Tyr Tyr Cys Ser 410  
 415  
 420

<210> 47  
 <211> 255  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc-feature  
 <223> Incyte ID No: 4184320CD1



<400> 47	Met Tyr Val Arg Val	Ser Phe Asp Thr Lys	Pro Asp Leu Leu Leu	15
	His Leu Met Thr Lys	Glu Trp Glu Leu Glu	Leu Pro Lys Leu Leu	30
	Ile Ser Val His Gly	Gly Leu Glu Asn Phe	Glu Leu Glu Pro Lys	45
	Leu Lys Glu Val Phe	Gly Lys Gly Leu Ile	Lys Ala Ala Met Thr	60
	Thr Gly Ala Trp Ile	Phe Thr Gly Gly Val	Asn Thr Gly Val Ile	75
	Arg His Val Gly Asp	Ala Leu Lys Asp His	Ala Ser Lys Ser Arg	90
	Gly Lys Ile Cys Thr	Ile Gly Ile Ala Pro	Trp Gly Ile Val Glu	105
	Asn Glu Glu Asp Leu	Ile Gly Arg Asp Val	Val Arg Pro Tyr Glu	120
	Thr Met Ser Asn Pro	Met Ser Lys Leu Thr	Val Leu Asn Ser Met	135
	His Ser His Phe Ile	Leu Ala Asp Asn Gly	Thr Thr Gly Lys Tyr	150
	Gly Ala Glu Val Lys	Leu Arg Arg Glu Leu	Glu Lys His Ile Ser	165
	Leu Glu Lys Ile Asn	Thr Arg Cys Leu Pro	Phe Phe Ser Leu Asp	180
	Ser Arg Leu Phe Tyr	Ser Phe Trp Gly Ser	Cys Glu Leu Asp Ser	195
	Val Gly Ile Gly Glu	Ile Ser Ile Val Leu	Glu Tyr Leu Arg Asp	210
	Gly Gly Pro Asn Val	Val Ile Val Val Val	Val Tyr Leu Arg Asp	225
	Thr Pro Pro Val Pro	Val Val Val Cys Asp	Gly Ser Gly Arg Ala	240
	Ser Asp Ile Leu Ala	Phe Gly His Lys Tyr	Ser Glu Glu Gly	255
<210> 48				
<211> 111				
<212> PRT				
<213> Homo sapiens				
<220>				
<221> misc_feature				
<223> Incyte ID No: 4764233CD1				
<400> 48	Met Ser Trp Arg Gly	Arg Ser Thr Tyr Arg	Pro Arg Pro Arg Arg	15
	Ser Leu Glu Pro Pro	Glu Leu Ile Gly Ala	Met Leu Glu Pro Thr	30
	Asp Glu Glu Pro Lys	Glu Glu Lys Pro Pro	Thr Lys Ser Arg Asn	45
	Pro Thr Pro Asp Glu	Lys Arg Glu Asp Asp	Gln Gly Ala Ala Glu	60
	Ile Glu Val Pro Asp	Leu Glu Ala Asp Leu	Gln Glu Leu Cys Glu	75
	Thr Lys Thr Gly Asp	Gly Cys Glu Gly Gly	Thr Asp Val Lys Gly	90
	Lys Ile Leu Pro Lys	Ala Glu His Phe Lys	Met Pro Glu Ala Gly	105
	Glu Gly Lys Ser Glu	Val		110
<210> 49				



<211> 422  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc-feature  
<223> Incyte ID No: 4817352CD1

<400> 49

Met Gly Lys Ala Lys Val Pro Ala Ser Lys Arg Ala Pro Ser Ser  
1  
Pro Val Ala Lys Pro Gly Pro Val Lys Thr Leu Thr Arg Lys Lys  
20  
Asn Lys Lys Lys Lys Lys Arg Phe Trp Lys Ser Lys Ala Arg Val  
35  
Ser Lys Lys Pro Ala Ser Gly Pro Gly Ala Val Val Arg Pro Pro  
50  
Lys Ala Pro Gln Asp Phe Ser Gln Asn Trp Lys Lys Ala Leu Gln Gln  
65  
Trp Leu Leu Lys Gln Lys Ser Gln Ala Pro Gln Lys Pro Leu Val  
80  
Ile Ser Gln Met Gly Ser Lys Lys Lys Pro Lys Ile Ile Gln Gln  
95  
Asn Lys Lys Gln Thr Ser Pro Gln Val Lys Gly Gln Gln Met Pro  
110  
Ala Gly Lys Asp Gln Ala Ser Arg Gly Ser Val Pro Ser Gly  
125  
Ser Lys Met Asp Arg Ala Pro Val Pro Arg Thr Lys Ala Ser  
140  
Gly Thr Gln His Asn Lys Lys Gly Thr Lys Gln Arg Thr Asn Gly  
155  
Asp Ile Val Pro Gln Arg Gly Asp Ile Gln His Lys Lys Arg Lys  
170  
Ala Lys Gln Ala Ala Pro Ala Pro Thr Gln Gln Asp Ile Trp  
185  
Phe Asp Asp Val Asp Pro Ala Asp Ile Gln Ala Ile Gly Pro  
200  
Gln Ala Ala Lys Ile Ala Arg Lys Gln Leu Gly Gln Ser Gln Gly  
215  
Ser Val Ser Leu Ser Leu Val Lys Gln Gln Ala Phe Gly Gly Leu  
230  
Thr Arg Ala Leu Ala Leu Asp Cys Gln Met Val Gly Val Gly Pro  
245  
Lys Gly Gln Gln Ser Met Ala Ala Arg Val Ser Ile Val Asn Gln  
260  
Val Gly Lys Cys Val Tyr Asp Lys Tyr Val Lys Pro Thr Gln Pro  
275  
Val Thr Asp Tyr Arg Thr Ala Val Ser Gly Ile Arg Pro Gln Asn  
290  
Leu Lys Gln Gly Gln Gln Leu Gln Val Val Gln Lys Gln Val Ala  
305  
Gln Met Leu Lys Gly Arg Ile Leu Val Gly His Ala Leu His Asn  
320  
Asp Leu Lys Val Leu Phe Leu Asp His Pro Lys Lys Lys Ile Arg  
335  
Asp Thr Gln Lys Tyr Lys Pro Phe Lys Ser Gln Val Lys Ser Gly  
350  
Arg Pro Ser Leu Arg Leu Leu Ser Gln Lys Ile Leu Gly Leu Gln  
365  
Val Gln Gln Ala Gln His Cys Ser Ile Gln Asp Ala Gln Ala Ala  
380  
Met Arg Leu Tyr Val Met Val Lys Lys Gln Trp Gln Ser Met Ala  
395





Arg Asp Arg Arg Pro Leu Leu Thr Ala Pro Asp His Cys Ser Asp  
410  
415  
420 Asp Ala

<210> 50  
<211> 397  
<212> PRT  
<213> Homo sapiens  
<220>  
<221> misc-feature  
<223> Incyte ID No: 5040573CD1  
<400> 50

Met Ala Met Ile Gln Leu Gly Phe Gly Arg Gln Asn Phe His Pro  
1  
5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000  
1005  
1010  
1015  
1020  
1025  
1030  
1035  
1040  
1045  
1050  
1055  
1060  
1065  
1070  
1075  
1080  
1085  
1090  
1095  
1100  
1105  
1110  
1115  
1120  
1125  
1130  
1135  
1140  
1145  
1150  
1155  
1160  
1165  
1170  
1175  
1180  
1185  
1190  
1195  
1200  
1205  
1210  
1215  
1220  
1225  
1230  
1235  
1240  
1245  
1250  
1255  
1260  
1265  
1270  
1275  
1280  
1285  
1290  
1295  
1300  
1305  
1310  
1315  
1320  
1325  
1330  
1335  
1340  
1345  
1350  
1355  
1360  
1365  
1370  
1375  
1380  
1385  
1390  
1395  
1400  
1405  
1410  
1415  
1420  
1425  
1430  
1435  
1440  
1445  
1450  
1455  
1460  
1465  
1470  
1475  
1480  
1485  
1490  
1495  
1500  
1505  
1510  
1515  
1520  
1525  
1530  
1535  
1540  
1545  
1550  
1555  
1560  
1565  
1570  
1575  
1580  
1585  
1590  
1595  
1600  
1605  
1610  
1615  
1620  
1625  
1630  
1635  
1640  
1645  
1650  
1655  
1660  
1665  
1670  
1675  
1680  
1685  
1690  
1695  
1700  
1705  
1710  
1715  
1720  
1725  
1730  
1735  
1740  
1745  
1750  
1755  
1760  
1765  
1770  
1775  
1780  
1785  
1790  
1795  
1800  
1805  
1810  
1815  
1820  
1825  
1830  
1835  
1840  
1845  
1850  
1855  
1860  
1865  
1870  
1875  
1880  
1885  
1890  
1895  
1900  
1905  
1910  
1915  
1920  
1925  
1930  
1935  
1940  
1945  
1950  
1955  
1960  
1965  
1970  
1975  
1980  
1985  
1990  
1995  
2000  
2005  
2010  
2015  
2020  
2025  
2030  
2035  
2040  
2045  
2050  
2055  
2060  
2065  
2070  
2075  
2080  
2085  
2090  
2095  
2100  
2105  
2110  
2115  
2120  
2125  
2130  
2135  
2140  
2145  
2150  
2155  
2160  
2165  
2170  
2175  
2180  
2185  
2190  
2195  
2200  
2205  
2210  
2215  
2220  
2225  
2230  
2235  
2240  
2245  
2250  
2255  
2260  
2265  
2270  
2275  
2280  
2285  
2290  
2295  
2300  
2305  
2310  
2315  
2320  
2325  
2330  
2335  
2340  
2345  
2350  
2355  
2360  
2365  
2370  
2375  
2380  
2385  
2390  
2395  
2400  
2405  
2410  
2415  
2420  
2425  
2430  
2435  
2440  
2445  
2450  
2455  
2460  
2465  
2470  
2475  
2480  
2485  
2490  
2495  
2500  
2505  
2510  
2515  
2520  
2525  
2530  
2535  
2540  
2545  
2550  
2555  
2560  
2565  
2570  
2575  
2580  
2585  
2590  
2595  
2600  
2605  
2610  
2615  
2620  
2625  
2630  
2635  
2640  
2645  
2650  
2655  
2660  
2665  
2670  
2675  
2680  
2685  
2690  
2695  
2700  
2705  
2710  
2715  
2720  
2725  
2730  
2735  
2740  
2745  
2750  
2755  
2760  
2765  
2770  
2775  
2780  
2785  
2790  
2795  
2800  
2805  
2810  
2815  
2820  
2825  
2830  
2835  
2840  
2845  
2850  
2855  
2860  
2865  
2870  
2875  
2880  
2885  
2890  
2895  
2900  
2905  
2910  
2915  
2920  
2925  
2930  
2935  
2940  
2945  
2950  
2955  
2960  
2965  
2970  
2975  
2980  
2985  
2990  
2995  
3000  
3005  
3010  
3015  
3020  
3025  
3030  
3035  
3040  
3045  
3050  
3055  
3060  
3065  
3070  
3075  
3080  
3085  
3090  
3095  
3100  
3105  
3110  
3115  
3120  
3125  
3130  
3135  
3140  
3145  
3150  
3155  
3160  
3165  
3170  
3175  
3180  
3185  
3190  
3195  
3200  
3205  
3210  
3215  
3220  
3225  
3230  
3235  
3240  
3245  
3250  
3255  
3260  
3265  
3270  
3275  
3280  
3285  
3290  
3295  
3300  
3305  
3310  
3315  
3320  
3325  
3330  
3335  
3340  
3345  
3350  
3355  
3360  
3365  
3370  
3375  
3380  
3385  
3390  
3395  
3400  
3405  
3410  
3415  
3420  
3425  
3430  
3435  
3440  
3445  
3450  
3455  
3460  
3465  
3470  
3475  
3480  
3485  
3490  
3495  
3500  
3505  
3510  
3515  
3520  
3525  
3530  
3535  
3540  
3545  
3550  
3555  
3560  
3565  
3570  
3575  
3580  
3585  
3590  
3595  
3600  
3605  
3610  
3615  
3620  
3625  
3630  
3635  
3640  
3645  
3650  
3655  
3660  
3665  
3670  
3675  
3680  
3685  
3690  
3695  
3700  
3705  
3710  
3715  
3720  
3725  
3730  
3735  
3740  
3745  
3750  
3755  
3760  
3765  
3770  
3775  
3780  
3785  
3790  
3795  
3800  
3805  
3810  
3815  
3820  
3825  
3830  
3835  
3840  
3845  
3850  
3855  
3860  
3865  
3870  
3875  
3880  
3885  
3890  
3895  
3900  
3905  
3910  
3915  
3920  
3925  
3930  
3935  
3940  
3945  
3950  
3955  
3960  
3965  
3970  
3975  
3980  
3985  
3990  
3995  
4000  
4005  
4010  
4015  
4020  
4025  
4030  
4035  
4040  
4045  
4050  
4055  
4060  
4065  
4070  
4075  
4080  
4085  
4090  
4095  
4100  
4105  
4110  
4115  
4120  
4125  
4130  
4135  
4140  
4145  
4150  
4155  
4160  
4165  
4170  
4175  
4180  
4185  
4190  
4195  
4200  
4205  
4210  
4215  
4220  
4225  
4230  
4235  
4240  
4245  
4250  
4255  
4260  
4265  
4270  
4275  
4280  
4285  
4290  
4295  
4300  
4305  
4310  
4315  
4320  
4325  
4330  
4335  
4340  
4345  
4350  
4355  
4360  
4365  
4370  
4375  
4380  
4385  
4390  
4395  
4400  
4405  
4410  
4415  
4420  
4425  
4430  
4435  
4440  
4445  
4450  
4455  
4460  
4465  
4470  
4475  
4480  
4485  
4490  
4495  
4500  
4505  
4510  
4515  
4520  
4525  
4530  
4535  
4540  
4545  
4550  
4555  
4560  
4565  
4570  
4575  
4580  
4585  
4590  
4595  
4600  
4605  
4610  
4615  
4620  
4625  
4630  
4635  
4640  
4645  
4650  
4655  
4660  
4665  
4670  
4675  
4680  
4685  
4690  
4695  
4700  
4705  
4710  
4715  
4720  
4725  
4730  
4735  
4740  
4745  
4750  
4755  
4760  
4765  
4770  
4775  
4780  
4785  
4790  
4795  
4800  
4805  
4810  
4815  
4820  
4825  
4830  
4835  
4840  
4845  
4850  
4855  
4860  
4865  
4870  
4875  
4880  
4885  
4890  
4895  
4900  
4905  
4910  
4915  
4920  
4925  
4930  
4935  
4940  
4945  
4950  
4955  
4960  
4965  
4970  
4975  
4980  
4985  
4990  
4995  
5000  
5005  
5010  
5015  
5020  
5025  
5030  
5035  
5040  
5045  
5050  
5055  
5060  
5065  
5070  
5075  
5080  
5085  
5090  
5095  
5100  
5105  
5110  
5115  
5120  
5125  
5130  
5135  
5140  
5145  
5150  
5155  
5160  
5165  
5170  
5175  
5180  
5185  
5190  
5195  
5200  
5205  
5210  
5215  
5220  
5225  
5230  
5235  
5240  
5245  
5250  
5255  
5260  
5265  
5270  
5275  
5280  
5285  
5290  
5295  
5300  
5305  
5310  
5315  
5320  
5325  
5330  
5335  
5340  
5345  
5350  
5355  
5360  
5365  
5370  
5375  
5380  
5385  
5390  
5395  
5400  
5405  
5410  
5415  
5420  
5425  
5430  
5435  
5440  
5445  
5450  
5455  
5460  
5465  
5470  
5475  
5480  
5485  
5490  
5495  
5500  
5505  
5510  
5515  
5520  
5525  
5530  
5535  
5540  
5545  
5550  
5555  
5560  
5565  
5570  
5575  
5580  
5585  
5590  
5595  
5600  
5605  
5610  
5615  
5620  
5625  
5630  
5635  
5640  
5645  
5650  
5655  
5660  
5665  
5670  
5675  
5680  
5685  
5690  
5695  
5700  
5705  
5710  
5715  
5720  
5725  
5730  
5735  
5740  
5745  
5750  
5755  
5760  
5765  
5770  
5775  
5780  
5785  
5790  
5795  
5800  
5805  
5810  
5815  
5820  
5825  
5830  
5835  
5840  
5845  
5850  
5855  
5860  
5865  
5870  
5875  
5880  
5885  
5890  
5895  
5900  
5905  
5910  
5915  
5920  
5925  
5930  
5935  
5940  
5945  
5950  
5955  
5960  
5965  
5970  
5975  
5980  
5985  
5990  
5995  
6000  
6005  
6010  
6015  
6020  
6025  
6030  
6035  
6040  
6045  
6050  
6055  
6060  
6065  
6070  
6075  
6080  
6085  
6090  
6095  
6100  
6105  
6110  
6115  
6120  
6125  
6130  
6135  
6140  
6145  
6150  
6155  
6160  
6165  
6170  
6175  
6180  
6185  
6190  
6195  
6200  
6205  
6210  
6215  
6220  
6225  
6230  
6235  
6240  
6245  
6250  
6255  
6260  
6265  
6270  
6275  
6280  
6285  
6290  
6295  
6300  
6305  
6310  
6315  
6320  
6325  
6330  
6335  
6340  
6345  
6350  
6355  
6360  
6365  
6370  
6375  
6380  
6385  
6390  
6395  
6400  
6405  
6410  
6415  
6420  
6425  
6430  
6435  
6440  
6445  
6450  
6455  
6460  
6465  
6470  
6475  
6480  
6485  
6490  
6495  
6500  
6505  
6510  
6515  
6520  
6525  
6530  
6535  
6540  
6545  
6550  
6555  
6560  
6565  
6570  
6575  
6580  
6585  
6590  
6595  
6600  
6605  
6610  
6615  
6620  
6625  
6630  
6635  
6640  
6645  
6650  
6655  
6660  
6665  
6670  
6675  
6680  
6685  
6690  
6695  
6700  
6705  
6710  
6715  
6720  
6725  
6730  
6735  
6740  
6745  
6750  
6755  
6760  
6765  
6770  
6775  
6780  
6785  
6790  
6795  
6800  
6805  
6810  
6815  
6820  
6825  
6830  
6835  
6840  
6845  
6850  
6855  
6860  
6865  
6870  
6875  
6880  
6885  
6890  
6895  
6900  
6905  
6910  
6915  
6920  
6925  
6930  
6935  
6940  
6945  
6950  
6955  
6960  
6965  
6970  
6975  
6980  
6985  
6990  
6995  
7000  
7005  
7010  
7015  
7020  
7025  
7030  
7035  
7040  
7045  
7050  
7055  
7060  
7065  
7070  
7075  
7080  
7085  
7090  
7095  
7100  
7105  
7110  
7115  
7120  
7125  
7130  
7135  
7140  
7145  
7150  
7155  
7160  
7165  
7170  
7175  
7180  
7185  
7190  
7195  
7200  
7205  
7210  
7215  
7220  
7225  
7230  
7235  
7240  
7245  
7250  
7255  
7260  
7265  
7270  
7275  
7280  
7285  
7290  
7295  
7300  
7305  
7310  
7315  
7320  
7325  
7330  
7335  
7340  
7345  
7350  
7355  
7360  
7365  
7370  
7375  
7380  
7385  
7390  
7395  
7400  
7405  
7410  
7415  
7420  
7425  
7430  
7435  
7440  
7445  
7450  
7455  
7460  
7465  
7470  
7475  
7480  
7485  
7490  
7495  
7500  
7505  
7510  
7515  
7520  
7525  
7530  
7535  
7540  
7545  
7550  
7555  
7560  
7565  
7570  
7575  
7580  
7585  
7590  
7595  
7600  
7605  
7610  
7615  
7620  
7625  
7630  
7635  
7640  
7645  
7650  
7655  
7660  
7665  
7670  
7675  
7680  
7685  
7690  
7695  
7700  
7705  
7710  
7715  
7720  
7725  
7730  
7735  
7740  
7745  
7750  
7755  
7760  
7765  
7770  
7775  
7780  
7785  
7790  
7795  
7800  
7805  
7810  
7815  
7820  
7825  
7830  
7835  
7840  
7845  
7850  
7855  
7860  
7865  
7870  
7875  
7880  
7885  
7890  
7895  
7900  
7905  
7910  
7915  
7920  
7925  
7930  
7935  
7940  
7945  
7950  
7955  
7960  
7965  
7970  
7975  
7980  
7985  
7990  
7995  
8000  
8005  
8010  
8015  
8020  
8025  
8030  
8035  
8040  
8045  
8050  
8055  
8060  
8065  
8070  
8075  
8080  
8085  
8090  
8095  
8100  
8105  
8110  
8115  
8120  
8125  
8130  
8135  
8140  
8145  
8150  
8155  
8160  
8165  
8170  
8175  
8180  
8185  
8190  
8195  
8200  
8205  
8210  
8215  
8220  
8225  
8230  
8235  
8240  
8245  
8250  
8255  
8260  
8265  
8270  
8275  
8280  
8285  
8290  
8295  
8300  
8305  
8310  
8315  
8320  
8325  
8330  
8335  
8340  
8345  
8350  
8355  
8360  
8365  
8370  
8375  
8380  
8385  
8390  
8395  
8400  
8405  
8410  
8415  
8420  
8425  
8430  
8435  
8440  
8445  
8450  
8455  
8460  
8465  
8470  
8475  
8480  
8485  
8490  
8495  
8500  
8505  
8510  
8515  
8520  
8525  
8530  
8535  
8540  
8545  
8550  
8555  
8560  
8565  
8570  
8575  
8580  
8585  
8590  
8595  
8600  
8605  
8610  
8615  
8620  
8625  
8630  
8635  
8640  
8645  
8650  
8655  
8660  
8665  
8670  
8675  
8680  
8685  
8690  
8695  
8700  
8705  
8710  
8715  
8720  
8725  
8730  
8735  
8740  
8745  
8750  
8755  
8760  
8765  
8770  
8775  
8780  
8785  
8790  
8795  
8800  
8805  
8810  
8815  
8820  
8825  
8830  
8835  
8840  
8845  
8850  
8855  
8860  
8865  
8870  
8875  
8880  
8885  
8890  
8895  
8900  
8905  
8910  
8915  
8920  
8925  
8930  
8935  
8940  
8945  
8950  
8955  
8960  
8965  
8970  
8975  
8980  
8985  
8990  
8995  
9000  
9005  
9010  
9015  
9020  
9025  
9030  
9035  
9040  
9045  
9050  
9055  
9060  
9065  
9070  
9075  
9080  
9085  
9090  
9095  
9100  
9105  
9110  
9115  
9120  
9125  
9130  
9135  
9140  
9145  
9150  
9155  
9160  
9165  
9170  
9175  
9180  
9185  
9190  
9195  
9200  
9205  
9210  
9215  
9220  
9225  
9230  
9235  
9240  
9245  
9250  
9255  
9260  
9265  
9270  
9275  
9280  
9285  
9290  
9295  
9300  
9305  
9310  
9315  
9320  
9325  
9330  
9335  
9340  
9345  
9350  
9355  
9360  
9365  
9370  
9375  
9380  
9385  
9390  
9395  
9400  
9405  
9410  
9415  
9420  
9425  
9430  
9435  
9440  
9445  
9450  
9455  
9460  
9465  
9470  
9475  
9480  
9485  
9490  
9495  
9500  
9505  
9510  
9515  
9520  
9525  
9530  
9535  
9540  
9545  
9550  
9555  
9560  
9565  
9570  
9575  
9580  
9585  
9590  
9595  
9600  
9605  
9610  
9615  
9620  
9625  
9630  
9635  
9640  
9645  
9650  
9655  
9660  
9665  
9670  
9675  
9680  
9685  
9690  
9695  
9700  
9705  
9710  
9715  
9720  
9725  
9730  
9735  
9740  
9745  
9750  
9755  
9760  
9765  
9770  
9775  
9780  
9785  
9790  
9795  
9800  
9805  
9810  
9815  
9820  
9825  
9830  
9835  
9840  
9845  
9850  
9855  
9860  
9865  
9870  
9875  
9880  
9885  
9890  
9895  
9900  
9905  
9910  
9915  
9920  
9925  
9930  
9935  
9940  
9945  
9950  
9955  
9960  
9965  
9970  
9975  
9980  
9985  
9990  
9995  
10000  
10005  
10010  
10015  
10020  
10025  
10030  
10035  
10040  
10045  
10050  
10055  
10060  
10065  
10070  
10075  
10080  
10085  
10090  
10095  
10100  
10105  
10110  
10115  
10120  
10125  
10130  
10135  
10140  
10145  
10150  
10155  
10160  
10165



365 His Phe Gly Leu Ala Ser Pro Phe Leu Ser Gly Leu Asn Leu 370  
 380 Leu Gly Lys Arg Lys Thr Arg 385  
 395

<210> 51  
 <211> 800  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Incyte ID No: 5627029CD1

<400> 51

Met Gly Ser Ser Lys Lys His Arg Gly Gly Lys Glu Ala Ala Gly  
 1  
 Thr Thr Ala Ala Ala Gly Thr Gly Gly Ala Thr Glu Gln Pro Pro  
 20  
 Arg His Arg Glu His Lys Lys His Lys His Arg Ser Gly Gly Ser  
 35  
 Gly Gly Ser Gly Gly Glu Arg Arg Lys Arg Ser Arg Gly Gly  
 50  
 Gly Glu Arg Gly Ser Gly Arg Arg Gly Ala Glu Ala Ala Arg  
 60  
 75  
 Ser Ser Thr His Gly Arg Glu Arg Ser Glu Ala Glu Pro Ser Glu  
 80  
 Arg Arg Val Lys Arg Glu Lys Arg Asp Asp Gly Tyr Glu Ala Ala  
 95  
 Ala Ser Ser Ser Lys Thr Ser Ser Gly Asp Ala Ser Ser Leu Ser Ile  
 110  
 Lys Lys Lys Lys Glu Ala Gly Thr Lys Glu Gln Pro  
 125  
 Glu Val Asn Ala Ile Lys Lys Glu Ala Gly Thr Lys Glu Gln Pro  
 140  
 Val Thr Ala Asp Val Ile Asn Pro Met Ala Leu Arg Gln Arg Glu  
 155  
 Glu Leu Arg Glu Lys Leu Ala Ala Lys Glu Lys Arg Leu Leu  
 170  
 Asn Gln Lys Leu Gly Lys Ile Lys Thr Leu Gly Glu Asp Asp Pro  
 185  
 Trp Leu Asp Asp Thr Ala Ala Trp Ile Glu Arg Ser Arg Gln Leu  
 200  
 Asp Lys Glu Lys Asp Leu Ala Glu Lys Arg Ala Lys Leu Leu Glu  
 215  
 Glu Met Asp Gln Phe Gly Val Ser Thr Leu Val Glu Gln Glu  
 230  
 Phe Gly Gln Arg Arg Gln Asp Leu Tyr Ser Ala Arg Asp Leu Gln  
 245  
 Gly Leu Thr Val Glu His Ala Ile Asp Ser Phe Arg Glu Gly Glu  
 260  
 Thr Met Ile Leu Thr Lys Asp Lys Val Leu Gln Glu Gln  
 275  
 Glu Asp Val Leu Val Asn Val Asn Leu Val Asp Lys Glu Arg Ala  
 290  
 Glu Lys Asn Val Glu Leu Arg Lys Lys Pro Asp Tyr Leu Pro  
 305  
 Tyr Ala Glu Asp Ser Val Asp Asp Leu Ala Glu Gln Lys Pro  
 320  
 Arg Ser Ile Leu Ser Lys Tyr Asp Glu Glu Gln Gly Glu Arg  
 335  
 Pro His Ser Phe Arg Leu Gln Gly Gly Thr Ala Asp Gly Leu  
 350



Arg	Glu	Arg	Glu	Leu	365
Ala	Gln	Ser	Thr	Val	370
Ala	Gln	Ser	Thr	Val	380
Leu	Thr	Pro	Glu	Met	385
Val	Thr	Thr	Val	Thr	390
Val	Thr	Thr	Val	Thr	400
Val	Thr	Thr	Val	Thr	405
Val	Thr	Thr	Val	Thr	410
Val	Thr	Thr	Val	Thr	415
Val	Thr	Thr	Val	Thr	420
Val	Thr	Thr	Val	Thr	425
Val	Thr	Thr	Val	Thr	430
Val	Thr	Thr	Val	Thr	435
Val	Thr	Thr	Val	Thr	440
Val	Thr	Thr	Val	Thr	445
Val	Thr	Thr	Val	Thr	450
Val	Thr	Thr	Val	Thr	455
Val	Thr	Thr	Val	Thr	460
Val	Thr	Thr	Val	Thr	465
Val	Thr	Thr	Val	Thr	470
Val	Thr	Thr	Val	Thr	475
Val	Thr	Thr	Val	Thr	480
Val	Thr	Thr	Val	Thr	485
Val	Thr	Thr	Val	Thr	490
Val	Thr	Thr	Val	Thr	495
Val	Thr	Thr	Val	Thr	500
Val	Thr	Thr	Val	Thr	505
Val	Thr	Thr	Val	Thr	510
Val	Thr	Thr	Val	Thr	515
Val	Thr	Thr	Val	Thr	520
Val	Thr	Thr	Val	Thr	525
Val	Thr	Thr	Val	Thr	530
Val	Thr	Thr	Val	Thr	535
Val	Thr	Thr	Val	Thr	540
Val	Thr	Thr	Val	Thr	545
Val	Thr	Thr	Val	Thr	550
Val	Thr	Thr	Val	Thr	555
Val	Thr	Thr	Val	Thr	560
Val	Thr	Thr	Val	Thr	565
Val	Thr	Thr	Val	Thr	570
Val	Thr	Thr	Val	Thr	575
Val	Thr	Thr	Val	Thr	580
Val	Thr	Thr	Val	Thr	585
Val	Thr	Thr	Val	Thr	590
Val	Thr	Thr	Val	Thr	595
Val	Thr	Thr	Val	Thr	600
Val	Thr	Thr	Val	Thr	605
Val	Thr	Thr	Val	Thr	610
Val	Thr	Thr	Val	Thr	615
Val	Thr	Thr	Val	Thr	620
Val	Thr	Thr	Val	Thr	625
Val	Thr	Thr	Val	Thr	630
Val	Thr	Thr	Val	Thr	635
Val	Thr	Thr	Val	Thr	640
Val	Thr	Thr	Val	Thr	645
Val	Thr	Thr	Val	Thr	650
Val	Thr	Thr	Val	Thr	655
Val	Thr	Thr	Val	Thr	660
Val	Thr	Thr	Val	Thr	665
Val	Thr	Thr	Val	Thr	670
Val	Thr	Thr	Val	Thr	675
Val	Thr	Thr	Val	Thr	680
Val	Thr	Thr	Val	Thr	685
Val	Thr	Thr	Val	Thr	690
Val	Thr	Thr	Val	Thr	695
Val	Thr	Thr	Val	Thr	700
Val	Thr	Thr	Val	Thr	705
Val	Thr	Thr	Val	Thr	710
Val	Thr	Thr	Val	Thr	715
Val	Thr	Thr	Val	Thr	720
Val	Thr	Thr	Val	Thr	725
Val	Thr	Thr	Val	Thr	730
Val	Thr	Thr	Val	Thr	735
Val	Thr	Thr	Val	Thr	740
Val	Thr	Thr	Val	Thr	745
Val	Thr	Thr	Val	Thr	750
Val	Thr	Thr	Val	Thr	755
Val	Thr	Thr	Val	Thr	760
Val	Thr	Thr	Val	Thr	765
Val	Thr	Thr	Val	Thr	770
Val	Thr	Thr	Val	Thr	775
Val	Thr	Thr	Val	Thr	780
Val	Thr	Thr	Val	Thr	785
Val	Thr	Thr	Val	Thr	790
Val	Thr	Thr	Val	Thr	795
Val	Thr	Thr	Val	Thr	800

<210> 52  
<211> 713  
<212> PRT



<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 5678487CD1

<400> 52

Met Ala Lys Ser Pro Glu Asn Ser Thr Leu Glu Ile Leu Gly  
1  
Gln Tyr Gln Arg Ser Leu Arg Glu His Ala Ser Arg Ser Ile His  
20  
Gln Leu Thr Cys Ala Leu Lys Glu Gly Asp Val Thr Ile Gln Ala  
35  
Asp Ala Pro Asn Leu Ser Phe Ser Thr Ser Val Gly Asn Glu Asp  
50  
Ala Arg Thr Ala Trp Pro Gln Leu Gln Gln Ser His Ala Val Asn  
65  
Gln Leu Lys Asp Leu Arg Gln Gln Ala Asp Lys Glu Ser Gln  
80  
Val Ser Pro Ser Arg Arg Lys Met Ser Pro Leu Arg Ser Leu  
95  
Glu His Glu Glu Thr Asn Met Pro Thr Met His Asp Leu Val His  
110  
Thr Ile Asn Asp Gln Ser Gln Tyr Ile His His Leu Glu Ala Glu  
125  
Val Lys Phe Cys Lys Glu Gln Leu Ser Gly Met Lys Asn Lys Ile  
140  
Gln Val Val Val Leu Asn Glu Gly Leu Gln Gln Gln Leu Lys  
155  
Ser Gln Arg Gln Glu Thr Leu Arg Glu Gln Thr Leu Leu Asp  
170  
Ala Ser Gly Asn Met His Asn Ser Trp Ile Thr Thr Gly Glu Asp  
185  
Ser Gly Val Gly Glu Thr Ser Lys Arg Pro Phe Ser His Asp Asn  
200  
Ala Asp Phe Gly Lys Ala Ala Ser Ala Gly Glu Gln Leu Glu Leu  
215  
Glu Lys Leu Lys Thr Tyr Glu Glu Lys Cys Glu Ile Glu Glu  
230  
Ser Gln Leu Lys Phe Leu Arg Asn Asp Leu Ala Glu Tyr Gln Arg  
245  
Thr Cys Glu Asp Leu Lys Glu Gln Leu Lys His Lys Glu Phe Leu  
260  
Leu Ala Ala Asn Thr Cys Asn Arg Val Gly Leu Cys Leu Lys  
275  
Cys Ala Gln His Glu Ala Val Leu Ser Thr His Thr Asn Val  
290  
His Met Gln Thr Ile Glu Arg Leu Val Lys Glu Arg Asp Leu  
305  
Met Ser Ala Leu Val Ser Val Arg Ser Ser Leu Ala Asp Thr Gln  
320  
Gln Arg Glu Ala Ser Ala Tyr Glu Gln Val Lys Gln Val Leu Gln  
335  
Ile Ser Glu Glu Ala Asn Phe Glu Lys Thr Lys Ala Leu Ile Gln  
350  
Cys Asp Gln Leu Arg Lys Glu Leu Arg Glu Ala Glu Arg Leu  
365  
Glu Lys Asp Leu Ala Ser Gln Gln Glu Lys Arg Ala Ile Glu Lys  
380  
Asp Met Met Lys Lys Glu Ile Thr Lys Glu Arg Glu Tyr Met Gly  
395  
Ser Lys Met Leu Ile Leu Ser Gln Asn Ile Ala Gln Leu Glu Ala  
410





Gln Val Gln Lys	Val Thr Lys Gln Lys	Ile Ser Ala Ile Asn Gln	425
Leu Gln Gln Ile	Gln Ser Gln Leu Ala	Ser Arg Gln Met Asp Val	435
Thr Lys Val Cys	Gly Gln Met Arg Tyr	Gln Leu Asn Lys Thr Asn	440
Met Gln Lys Asp	Gln Ala Gln Lys Gln	His Arg Gln Phe Arg Ala	445
Lys Thr Asn Arg	Asp Leu Gln Ile Lys	Asp Gln Gln Ile Gln Lys	450
Leu Arg Ile Gln	Leu Asp Gln Ser Lys	Gln His Leu Gln Gln	455
Gln Ala Leu Gln	Ala Gln Arg Gln Gln	Gln Thr Gln Lys	460
Ile Gln Gln Met	Gln Ala Gln His Asp	Lys Thr Gln Asn Gln	465
Thr Leu Leu Leu	Thr Phe Leu Thr Lys	Leu Lys Leu Lys	470
Gln Gln Cys Cys	Thr Leu Ala Lys Lys	Gln Gln Ile Ser	475
Lys Thr Arg Ser	Gln Ile Ala Gln Leu	Ser Gln His Gln	480
Thr Tyr Asp Lys	Leu Gly Lys Leu Gln	Arg Asn Gln Gln	485
Gln Gln Gln Cys	Val Gln His Gly Arg	Val His Gln Thr Met Lys	490
Gln Arg Leu Arg	Gln Leu Asp Lys His	Ser Gln Ala Thr Ala Gln	495
Gln Leu Val Gln	Leu Ser Lys Gln Asn	Gln Leu Leu Leu Gln	500
Arg Gln Ser Leu	Ser Gln Gln Val Asp	Arg Leu Arg Thr Gln	505
Pro Ser Met Pro	Gln Ser Asp Cys		510
<210> 53			
<211> 880			
<212> PRT			
<213> Homo sapiens			
<220>			
<221> misc-feature			
<223> Incyte ID No: 5682976CD1			
<400> 53			
Met Ser Arg Gly	Ser Cys Pro His	Leu Trp Asp Val Arg	1
Lys Arg Ser Leu	Gly Leu Gln Asp Pro	Ser Arg Leu Arg Ser	15
Tyr Leu Gly Arg	Gln Phe Ile Gln	Arg Leu Lys Leu Gln	30
Thr Leu Asn Val	His Asp Gly Cys	Val Asn Thr Ile Cys	45
Asp Thr Gly Tyr	Ile Leu Ser Gly	Asp Asp Thr Lys Leu	60
Val Ile Ser Asn	Pro Tyr Ser Arg Lys	Val Leu Thr Thr Ile	75
Ser Gly His Arg	Ala Asn Ile Phe	Ser Ala Lys Phe Leu	90
		Pro Cys	90



95	Thr Asn Asp Lys Gln Ile Val Ser Cys Ser Gly Asp Gly Val Ile	100	Thr Asn Asp Lys Gln Ile Val Ser Cys Ser Gly Asp Gly Val Ile
110	Phe Tyr Thr Asn Val Gln Gln Asp Ala Gln Thr Asn Arg Gln Cys	115	Phe Tyr Thr Asn Val Gln Gln Asp Ala Gln Thr Asn Arg Gln Cys
125	Gln Phe Thr Cys His Tyr Gly Thr Thr Tyr Gln Ile Met Thr Val	130	Gln Phe Thr Cys His Tyr Gly Thr Thr Tyr Gln Ile Met Thr Val
140	Pro Asn Asp Pro Tyr Phe Leu Ser Cys Gly Gln Asp Gly Thr	145	Pro Asn Asp Pro Tyr Phe Leu Ser Cys Gly Gln Asp Gly Thr
155	Val Arg Trp Phe Asp Thr Arg Ile Lys Thr Ser Cys Thr Lys Gln	160	Val Arg Trp Phe Asp Thr Arg Ile Lys Thr Ser Cys Thr Lys Gln
170	Asp Cys Lys Asp Ile Leu Ile Asn Cys Arg Arg Ala Ala Thr	175	Asp Cys Lys Asp Ile Leu Ile Asn Cys Arg Arg Ala Ala Thr
185	Ser Val Ala Ile Cys Pro Ile Pro Tyr Tyr Leu Ala Val Gly	190	Ser Val Ala Ile Cys Pro Ile Pro Tyr Tyr Leu Ala Val Gly
200	Cys Ser Asp Ser Val Arg Ile Tyr Asp Arg Arg Met Leu Gly	205	Cys Ser Asp Ser Val Arg Ile Tyr Asp Arg Arg Met Leu Gly
215	Thr Arg Ala Thr Gly Asn Tyr Ala Gly Arg Gly Thr Thr Gly Met	220	Thr Arg Ala Thr Gly Asn Tyr Ala Gly Arg Gly Thr Thr Gly Met
230	Val Ala Arg Phe Ile Pro Ser His Leu Asn Lys Ser Cys Arg	235	Val Ala Arg Phe Ile Pro Ser His Leu Asn Lys Ser Cys Arg
245	Val Thr Ser Leu Cys Tyr Ser Gln Asp Gly Gln Gln Ile Leu Val	250	Val Thr Ser Leu Cys Tyr Ser Gln Asp Gly Gln Gln Ile Leu Val
260	Ser Tyr Ser Ser Asp Tyr Ile Tyr Leu Phe Asp Pro Lys Asp	265	Ser Tyr Ser Ser Asp Tyr Ile Tyr Leu Phe Asp Pro Lys Asp
275	Thr Ala Arg Gln Leu Lys Thr Pro Ser Ala Gln Gln Arg Arg Gln	280	Thr Ala Arg Gln Leu Lys Thr Pro Ser Ala Gln Gln Arg Arg Gln
290	Glu Leu Arg Gln Pro Val Lys Arg Leu Arg Arg Gly Asp	295	Glu Leu Arg Gln Pro Val Lys Arg Leu Arg Arg Gly Asp
305	Pro Val Lys Arg Leu Arg Arg Leu Arg Arg Gly Asp	310	Pro Val Lys Arg Leu Arg Arg Leu Arg Arg Gly Asp
320	Trp Ser Asp Thr Gly Pro Arg Ala Arg Pro Gln Ser Gln Arg	325	Trp Ser Asp Thr Gly Pro Arg Ala Arg Pro Gln Ser Gln Arg
335	Arg Asp Gly Gln Gln Ser Pro Asn Val Ser Leu Met Gln Arg Met	340	Arg Asp Gly Gln Gln Ser Pro Asn Val Ser Leu Met Gln Arg Met
350	Ser Asp Met Leu Ser Arg Trp Phe Gln Ala Ser Gln Val Val	355	Ser Asp Met Leu Ser Arg Trp Phe Gln Ala Ser Gln Val Val
365	Gln Ser Asn Arg Gly Arg Gly Arg Ser Arg Pro Arg Gly Thr	370	Gln Ser Asn Arg Gly Arg Gly Arg Ser Arg Pro Arg Gly Thr
380	Ser Gln Ser Asp Ile Ser Thr Leu Pro Thr Val Pro Ser Ser Pro	385	Ser Gln Ser Asp Ile Ser Thr Leu Pro Thr Val Pro Ser Ser Pro
395	Asp Leu Gln Val Ser Gln Thr Ala Met Gln Val Asp Thr Pro Ala	400	Asp Leu Gln Val Ser Gln Thr Ala Met Gln Val Asp Thr Pro Ala
410	Gln Gln Phe Leu Gln Pro Ser Thr Ser Ser Thr Met Ser Ala Gln	415	Gln Gln Phe Leu Gln Pro Ser Thr Ser Ser Thr Met Ser Ala Gln
425	Ala His Ser Thr Ser Pro Thr Gln Ser Pro His Ser Thr Pro	430	Ala His Ser Thr Ser Pro Thr Gln Ser Pro His Ser Thr Pro
440	Leu Leu Ser Ser Pro Asp Ser Gln Gln Arg Gln Ser Val Gln Ala	445	Leu Leu Ser Ser Pro Asp Ser Gln Gln Arg Gln Ser Val Gln Ala
455	Ser Gly His His Thr His Gln Ser Asp Ser Pro Ser Ser Val	460	Ser Gly His His Thr His Gln Ser Asp Ser Pro Ser Ser Val
470	Val Asn Lys Gln Leu Gly Ser Met Ser Leu Asp Gln Gln Asp	475	Val Asn Lys Gln Leu Gly Ser Met Ser Leu Asp Gln Gln Asp
485	Asn Asn Asn Gln Lys Leu Ser Pro Lys Gly Thr Gly Gln Pro	490	Asn Asn Asn Gln Lys Leu Ser Pro Lys Gly Thr Gly Gln Pro
500	Val Leu Ser Leu His Tyr Ser Thr Gln Gly Thr Thr Thr Ser Thr	505	Val Leu Ser Leu His Tyr Ser Thr Gln Gly Thr Thr Thr Ser Thr
515	Ile Lys Leu Asn Phe Thr Asp Gln Trp Ser Ser Ile Ala Ser Ser	520	Ile Lys Leu Asn Phe Thr Asp Gln Trp Ser Ser Ile Ala Ser Ser
530	Ser Arg Gly Ile Gly Ser His Cys Lys Ser Gln Gly Gln Gln	535	Ser Arg Gly Ile Gly Ser His Cys Lys Ser Gln Gly Gln Gln
545	Ser Phe Val Pro Gln Ser Ser Val Gln Pro Gln Gly Asp Ser	550	Ser Phe Val Pro Gln Ser Ser Val Gln Pro Gln Gly Asp Ser
560	Gln Thr Lys Ala Pro Gln Ser Ser Gln Ser Ser Gln Asp Val Thr Lys	565	Gln Thr Lys Ala Pro Gln Ser Ser Gln Ser Ser Gln Asp Val Thr Lys



Gln Gln Gly Val Ser Ala Gln Asn Pro Val Gln Asn His Ile Asn  
 575 580 585  
 Ile Thr Gln Ser Asp Lys Phe Thr Ala Lys Pro Leu Asp Ser Asn  
 590 595  
 Ser Gly Gln Arg Asn Asp Leu Asn Leu Asp Arg Ser Cys Gly Val  
 600 605 610 615  
 Pro Gln Gln Ser Ala Ser Ser Gln Lys Ala Lys Gln Pro Gln Thr  
 620 625 630  
 Ser Asp Gln Thr Ser Thr Gln Ser Ala Thr Asn Gln Asn Asn Thr  
 635 640 645  
 Asn Pro Gln Pro Gln Phe Gln Thr Gln Ala Thr Gly Pro Ser Ala  
 650 655 660  
 His Gln Gln Thr Ser Thr Arg Asp Ser Ala Leu Gln Asp Thr Asp  
 665 670 675  
 Asp Ser Asp Asp Asp Pro Val Leu Ile Pro Gly Ala Arg Tyr Arg  
 680 685 690 705  
 Ala Gly Pro Gly Arg Arg Ser Ala Val Ala Arg Ile Gln Gln  
 710 715 720 725  
 Phe Phe Arg Arg Arg Lys Gln Arg Lys Gln Met Gln Gln Leu Asp  
 730 735 740 745 750  
 Thr Leu Asn Ile Arg Arg Pro Leu Val Lys Met Val Tyr Lys Gly  
 755 760 765 770 775 780 785 790 795  
 His Arg Asp Arg His Thr Ala Gln His Leu Met Leu Gln Ala Asp  
 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875  
 Gly Gln Gln Asn Gln Asp Gln Gln

&lt;210&gt; 54

&lt;211&gt; 855

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5992432CD1

&lt;400&gt; 54

Met Val Val Met Ala Arg Leu Ser Arg Pro Gln Arg Pro Asp Leu  
 1 5 10 15  
 Val Phe Gln Gln Asp Leu Pro Tyr Gln Gln Ile Met Arg  
 20 25 30  
 Asn Gln Phe Ser Val Lys Cys Trp Leu Arg Tyr Ile Gln Phe Lys  
 35 40 45  
 Gln Gly Ala Pro Lys Pro Arg Leu Asn Gln Leu Tyr Gln Arg Ala  
 50 55 60  
 Leu Lys Leu Leu Pro Cys Ser Tyr Lys Leu Trp Tyr Arg Tyr Leu  
 65 70 75  
 Lys Ala Arg Arg Ala Gln Val Lys His Arg Cys Val Thr Asp Pro



90	85	80
Ala Tyr Glu Asp	Val Asn Cys His	Val Asn Cys His
105	100	95
Met His Lys Met	Pro Arg Leu Trp Leu	Asp Tyr Cys Gln Phe
120	115	110
Met Asp Arg	Val Thr His Thr	Arg Thr Phe Asp
135	130	125
Ala Leu Arg Ala	Leu Pro Ile Thr	Gln His Ser Arg Ile
140	145	140
Leu Tyr Leu Arg	Phe Leu Arg Ser His	Pro Leu Pro Gln Thr
155	160	165
Val Arg Gly Tyr	Arg Phe Leu Lys	Leu Ser Pro Gln Ser
170	175	180
Glu Glu Tyr Ile	Tyr Leu Lys Ser	Asp Arg Leu Asp
185	190	195
Ala Ala Gln Arg	Leu Val Val Asn	Asp Gln Arg Phe
200	205	210
Ser Lys Ala Gly	Ser Asn Tyr Gln	Leu Trp His Gln Cys
215	220	225
Asp Leu Ile Ser	Gln Ser Leu Asn Val	Leu Thr Asp Gln
230	235	240
Asp Ala Ile Ile	Arg Gly Leu Thr	Arg Phe Thr Asn
245	250	255
Gly Lys Leu Trp	Cys Ser Leu Ala Asp	Tyr Ile Arg Ser
260	265	270
His Phe Glu Lys	Ala Arg Asp Val	Gln Ala Ile Arg
275	280	285
Val Met Thr Val	Arg Asp Phe Thr	Gln Val Tyr Ala
290	295	300
Gln Phe Glu Glu	Ser Met Ile Ala	Lys Met Gln Thr
305	310	315
Glu Leu Gly Arg	Glu Asp Val	Asp Leu Glu Arg
320	325	330
Leu Ala Arg Phe	Gln Leu Ile Ser	Arg Arg Pro Leu
335	340	345
Asn Ser Val Leu	Arg Gln Asn Pro	His Val His Gln
350	355	360
His Lys Arg Val	Ala Leu His Gln	Arg Pro Arg Glu
365	370	375
Asn Thr Tyr Thr	Glu Val Gln Thr	Val Asp Pro Phe
380	385	390
Thr Gly Lys Pro	His Thr Leu Trp	Val Ala Lys Phe
395	400	405
Glu Asp Asn Gly	Gln Leu Asp Ala	Arg Val Ile Leu
410	415	420
Ala Thr Lys Val	Asn Phe Lys Gln	Val Asp Leu Ala
425	430	435
Trp Cys Gln Cys	Gly Gln Leu Glu	Arg His Gln Asn
440	445	450
Glu Ala Leu Arg	Leu Arg Lys Ala	Thr Ala Leu Pro
455	460	465
Arg Ala Glu Tyr	Phe Asp Gly Ser	Gln Asn Arg Val
470	475	480
Tyr Lys Ser Leu	Lys Val Trp Ser	Met Leu Ala Asp
485	490	495
Ser Leu Gly Thr	Phe Gln Ser Thr	Lys Ala Val Tyr
500	505	510
Leu Asp Leu Arg	Ile Ala Thr Pro	Gln Ile Asn Tyr
515	520	525
Met Phe Leu Glu	Gln His Lys Tyr	Phe Lys Ala
530	535	540
Tyr Glu Arg Gly	Ile Ser Leu Phe	Lys Trp Pro Asn
545	550	555





Ile Trp Ser Thr Tyr Leu Thr Lys Phe Ile Ala Arg Tyr Gly 560  
 Arg Lys Leu Gln Arg Ala Arg Asp Leu Phe Gln Ala Leu Asp 570  
 Gly Cys Pro Pro Lys Tyr Ala Lys Thr Leu Tyr Leu Tyr Ala 580  
 Gln Leu Gln Gln Gln Trp Gly Leu Ala Arg His Ala Met Ala Val 590  
 Tyr Gln Arg Ala Thr Arg Ala Val Gln Pro Ala Gln Gln Tyr Asp 600  
 Met Phe Asn Ile Tyr Ile Lys Arg Ala Gln Ile Tyr Gly Val 610  
 Thr His Thr Arg Gly Ile Tyr Gln Lys Ala Ile Gln Val Leu Ser 620  
 Asp Gln His Ala Arg Gln Met Cys Leu Arg Phe Ala Asp Met Gln 630  
 Cys Lys Leu Gly Ile Asp Arg Ala Arg Ala Ile Tyr Ser Phe 640  
 Cys Ser Gln Ile Cys Asp Pro Arg Thr Thr Gly Ala Phe Trp Gln 650  
 Thr Trp Lys Asp Phe Gln Val Arg His Gly Asn Gln Asp Thr Ile 660  
 Lys Gln Met Leu Arg Ile Arg Arg Ser Val Gln Ala Thr Tyr Asn 670  
 Ser Ala Thr Gly Thr Val Ser Asp Leu Ala Pro Gly Gln Ser Gly 680  
 Met Asp Asp Met Lys Leu Leu Gln Arg Ala Gln Gln Leu Ala 690  
 Ala Gln Ala Gln Arg Asp Gln Pro Leu Arg Ala Gln Ser Lys Ile 700  
 Leu Phe Val Arg Ser Asp Ala Ser Arg Gln Leu Ala Gln Leu 710  
 Thr Gln Val Asn Phe Ala Ser Gln Met Leu Lys Val Ser Gly 720  
 Thr Trp Lys Asp Phe Gln Val Arg His Gly Asn Gln Asp Thr Ile 730  
 Lys Gln Met Leu Arg Ile Arg Arg Ser Val Gln Ala Thr Tyr Asn 740  
 Asp Gln His Ala Arg Gln Met Cys Leu Arg Phe Ala Asp Met Gln 750  
 Cys Lys Leu Gly Ile Asp Arg Ala Arg Ala Ile Tyr Ser Phe 760  
 Tyr Gln Arg Gly Ile Tyr Gln Lys Ala Ile Gln Val Leu Ser 770  
 Thr Trp Lys Asp Phe Gln Val Arg His Gly Asn Gln Asp Thr Ile 780  
 Ala Gln Ala Gln Arg Asp Gln Pro Leu Arg Ala Gln Ser Lys Ile 790  
 Leu Phe Val Arg Ser Asp Ala Ser Arg Gln Leu Ala Gln Leu 800  
 Ala Gln Val Asn Pro Gln Ile Gln Leu Gly Gln Asp Gln 810  
 Asp Gln Ser Val Pro Ala Val Phe Gly Ser Leu Lys Gln Asp 820  
 Met Asp Asp Met Lys Leu Leu Gln Arg Ala Gln Gln Leu Ala 830  
 Gln Gln Ser Val Pro Ala Val Phe Gly Ser Leu Lys Gln Asp 840  
 Gln Gln Ser Val Pro Ala Val Phe Gly Ser Leu Lys Gln Asp 850

<210> 55

<211> 1598

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Inyte ID No: 116462CBI

<400> 55

atttactatgt aacttgcccc ttgtctcaact catgtccttgcc ttgtggaacc gtttgaagaac 120  
 tgaatatccat tggcgggttagg aaccacagatc cccgggcattc ccagtgtctcc gagtcccttcg 180  
 ggccttccttc tccgggtcttc gagggtcttcg aaacccgaaac cgtctgtctcg tgggcgcagc 240  
 gccgagatcg atctcaacttc acctgtgtcg cactccagct gaccccaagta ggaagccaga 300  
 cgaagctgttaa aacatgaaacg gaagagtgga ttatttggct actgaggaag agatccaatct 360  
 taccaagaggc cccctcagggc tgggcttccaa catcgtcgggt gggaacagatc agcagttatgt 420  
 ctccaagac agtggcatct acgtcagccg catccaagaag aatggggctcg cggcccttga 480  
 tgggcggctc caggaggggtg ataagatcct ttcggttaat ggccaagacc taaagaacct 540  
 gctgcacccag gatgtcttag acctcttcg taatgcagggc tatgtctgtc ctctgagagt 600  
 gcagcacacag ttacaggtcg agaatggacc tataggaacat cgaagttgaag gggaccccaag 660  
 tgttatcccc atattatgg tgcgtgtgcc agtgtttgcc ctcaaccatgg tagcagccctg 720



gactttcag agataccggc aacaacttgc aaaaacttgc tctcttccaa tactcccaat 780  
gaagatatac tccaccctcc gctactctgc catgctcttc cctcctctg 840  
catagccaga tttagaagtga ctgtatacca ccccaaacct tgtgttcaac agtctccaat 900  
tctctatat ctaattgggaa agtlaaagtga ttgtttgaa gaaaacctgaa gaaagaactc 960  
ggctctatac aaatgaggaag tcatatactc tactagagaa ttctagctac attccagctac 1020  
gaacccaaga gagaagaatc gactcttccc gtcacccatag gcaataacct ttctcttagc 1080  
tggcatggcca taagggccag ctatgtgata tttagaggaag aaaagatttc tctttttaat 1140  
gactctcttc gggaatatc ttgtggcttc atttaattcc taacctacgta ccttggtgccc 1200  
tatatcgaca aagaagtgaag agagcatctc tacttttca aaaaagcaaa tactatat 1260  
cacatacgta tgcataatc atagtatac atgtatcccc atgtagaaat aagagtgaag 1320  
aagctactca gttctctgaag aaaaactga aatacaactg atcttaacct aggaattcaa 1380  
agagatatc aacaagagca aatgtatag gaaacaaggga gttgagccta taagggagc 1440  
ttcagttcca ttatataaa taataactga gactgggaga ttgtctgagt agcctgggca 1500  
cccaagcact tggtagcccg aagtgggaga ttgtctgagt ccaaggagacc agcctgggca 1560  
accatagcaaa accctactc tacttaaaaa aaaaaaaa

&lt;210&gt; 56

&lt;211&gt; 1432

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Inctye ID No: 1210462CB1

&lt;400&gt; 56

tgagaatgaa agtggatgac cgcgaatccc ggaagtcaaa ctgtttttt cagtccctg 60  
gaggtcttct gatatactgat cgcgtacacc ttgtgttca aagctctcag aagctctcag 120  
ctgaacccaaa atcagaaggagg aagctctgcaa aacagatttg caaagtctg 180  
ttgatatcata ttgaaaacaa gtaatcccaa gaactcggag atgcctggaa tacagttaag 240  
gagatatcata catctccatc atgctggcaa tatgctggcc tgcctaaccc agttcaattat 300  
cccttttgaa cttgaataagg tttaacttgc aagggtctatc accaactctc tcaagggtact 360  
tcaaccatc atcttaaatc agtgaagtgc taacctagca gaaactcggg ccgaatccct 420  
tcaagaatg aaccatlgaa aatatatcac tccaaatgc tgcctctcct tgcctctcct 480  
cctccagttg atcgaaggga ttgaaggagg ttgtgcttgc caggttatct tcatgttaat 600  
gaaatatgata gttcgtgact gaagtgtgctc aggcagacatc tcatgttaat 660  
cccttggata atgttatcaa agtgtcttga ttggatlgca gaaaaattgg agtatgcacag 720  
ccttgaatgt ttgaacaagt tttagttgat gctccgtgtc caaatgtatcg aagctgtgtg 780  
tttctctcgt actctcagaa ggcatactct aagctatc aaagggtgaa ttgtccctct 840  
ctacagtaag gtcctgtcaa agtctgtcaa aatcaagatg tgatccagtga aattttaaac 960  
tactctacat gcaagccttc caagcagaa aatcaagatg tgatccagtga aattttaaac 960  
tcccacgcta acatcatgccc taaggatatg caaaggatatg ctaaggtactg caatttaaac 1020  
ttccatctcgt aactcagaaa aagatlaaagc aaggtatgaa ttgtccctct 1080  
ccttgaatgt aatgaatctc atgtatgata ttctagagaa ttctagctac attccagctac 1140  
ggctctatac aaatgaggaag tcatatactc tactagagaa ttctagctac attccagctac 1200  
tatatcgaca aagaagtgaag agagcatctc tacttttca aaaaagcaaa tactatat 1260  
cacatacgta tgcataatc atagtatac atgtatcccc atgtagaaat aagagtgaag 1320  
aagctactca gttctctgaag aaaaactga aatacaactg atcttaacct aggaattcaa 1380  
agagatatc aacaagagca aatgtatag gaaacaaggga gttgagccta taagggagc 1440  
ttcagttcca ttatataaa taataactga gactgggaga ttgtctgagt agcctgggca 1500  
cccaagcact tggtagcccg aagtgggaga ttgtctgagt ccaaggagacc agcctgggca 1560  
accatagcaaa accctactc tacttaaaaa aaaaaaaa ataatctaga atagtgtgtc 1620  
aataatctag caatctagaa tctaggtctc agaatctgac gttgtgtctc attcaataac 1680  
gactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 1740  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 1800  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 1860  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 1920  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 1980  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2040  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2100  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2160  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2220  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2280  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2340  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2400  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2460  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2520  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2580  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2640  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2700  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2760  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2820  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2880  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 2940  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3000  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3060  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3120  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3180  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3240  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3300  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3360  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3420  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3480  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3540  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3600  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3660  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3720  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3780  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3840  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3900  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 3960  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4020  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4080  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4140  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4200  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4260  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4320  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4380  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4440  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4500  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4560  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4620  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4680  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4740  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4800  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4860  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4920  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 4980  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5040  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5100  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5160  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5220  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5280  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5340  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5400  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5460  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5520  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5580  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5640  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5700  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5760  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5820  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5880  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 5940  
tactctctcgt aactcagaaa aagctctcag aatcaagatg tgatccagtga aattttaaac 6000

&lt;400&gt; 57

gaggtgtcag ctacgggagg cgcacatctg gaggggacaa aactccggcg acacgagtgaa 60  
cacaaataaa cccctggacc cctctgttcc ctcaagctca agggccggga tgttgtacct 120



58	<210>
1774	<211>
DNA	<212>
HOMO	<213>

<213> Homo sapiens

```
<220>  
<221> misc_feature  
<223> Incyte ID No: 1416289CB1
```

[illegible]



aacttgaca gcaagcttct gactctccct tcatatctc gctgtaacct cttgtatata 960  
tcaaccaaaa aaaaagattg aaaaataatcg tcaaccagaa aatccagaaa actggaagt 1020  
tcaatcagtag gaaacagtag caactttagaa aactttagaa ccaagctttaa tttaattgac 1080  
ctactgatat tcaatcagaa ggtgactaaag aatgacaagaa gcttatgaa gcttatgaa 1140  
aatataagaa atatatctta tccatcatgca atttcatgca atttcatgca atttcatgca 1200  
aagccaagaa aatatctgct aaaaactatct tgttagaaatc atgttagaaatc atgttagaaatc 1260  
tttagcatca atagaaaatc gctgttagaa aatctcaatc tcatctgcaaa caaataatag 1320  
atttaattt tagcttaaac tttgttctca ccttatgtta gttgaaacctca gttatccatc 1380  
tgtaaatctc ttttatattg gctaaaataa tctaaaagaa taaatttgggt ggccaattag 1440  
aaatgctctc ttcaagttagt gtaattgaaag cttctctca acaattctca cttctctca 1500  
tgattctcc ttttagtcta atactctcc aggtcatatc tgttttaat catataat 1560  
tttctctcg gttttagaa ctaagcttagt aaaaacttct taaaacttaa gcatgttcat 1620  
tgctattct ttaatttga cttcttag agtttagaa cagtttagaa ccaaggtatc tttctgctc 1680  
cctcaatg atcttgaa agggaaaatc agtttagaa cagtttagaa ccaaggtatc tttctgctc 1740  
acatgctct tttcttgaa gttcttgaca gttctgtacac cttc 1774

<210> 59  
<211> 1268  
<212> DNA  
<213> Homo sapiens  
<220>  
<221> misc\_feature  
<223> Inyte ID No: 1558289CB1

<400> 59

taagtgaagc ttctccatc tcttagctc ccggtgaacat ccaagycaaag aactgaccc 60  
agcacagcaa tgaatcccaa catataccaa tcaatgagaa ccaatgagaa ccaatgagaa 120  
agaaagcagaa ctaactgagaa agaaagcagaa agaaagcagaa agaaagcagaa agaaagcagaa 180  
gactgaccca gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca 240  
gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca 300  
gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca 360  
gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca 420  
gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca 480  
gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca 540  
tttgatcata tctctgaccc aggtctgctc aagttctgctc aagttctgctc aagttctgctc 600  
caacttgcaa agatctgtag cgtgactcag aagttctgtag aagttctgtag aagttctgtag 660  
ccttgagcct agcatatc aacttgacaa aatggttagc ttgagtttagc ttgagtttagc 720  
ccttgagcct agcatatc aacttgacaa aatggttagc ttgagtttagc ttgagtttagc 780  
gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca gctgaccca 840  
caagagtgct tttagactg agaacagccc aacctccaaa agcttagttaga gaaagtagcaa 900  
tgcctgaggg ttgagcctc tctccatccc agtccccagaa caggtaaaacaa aactgcttaga 960  
aaaaagttgaa gttgaaacct gttgaaacct gttgaaacct gttgaaacct gttgaaacct 1020  
gccccccc atgtgaaacca agtgaagagaa gtaagaccca gtaagaccca gtaagaccca 1080  
ttaaagacaa agaggtgagaa gttgcccctta gtagaacagagaa gtagaacagagaa gtagaacagagaa 1140  
cgttcccaa gttgagtagaa aaggtgtaga ggtgagtaga ggtgagtaga ggtgagtaga 1200  
actactaaa tatagaaaac aaaaagagagc ccaagtagagc gcaaacagagc gcaaacagagc 1260  
aaggaaaagaa gggagagagc gggagagagc gggagagagc gggagagagc gggagagagc 1268

<210> 60

<211> 1331  
<212> DNA  
<213> Homo sapiens  
<220>  
<221> misc\_feature  
<223> Inyte ID No: 1577739CB1

<400> 60

gacatcttga ggcacagctg gacctgagtc cttgtgtcagaa ctttaatacaa cttgaagagtc 60  
cagccacccc gttgagagcc ttgagctgaa ttgtgaaagaa gttgattgagc ttgagcttctca 120  
gctgttgaag gccgagatgg aaaaagtctt ttgcaagaaacaa gtagctcttca gtagcttctca 180  
gaaagggaaag atttgcctgc gctgcccggc caaagtccccg cttgttctcgt ggtccgcccag 240





ctgtctcttc tgcagaagag cgtctcgac tctctgtagc ataaagatga agatgcttc 300  
 taagaattt ggaacatctc ctgtctacac actggcttt gagagcttc agatgcttc 360  
 agctggcaaa accggccaa tccagagaag agacatctt cagtcttc aagggccaa 420  
 gttggcagag gttggagag cgttcccca catctacc caagctgtg tccagaggt 480  
 tgtctcaggt gattggacca gcttgttg agacatctc cagtgcttg tccagaggt 540  
 ggaagcttc aacatcagc cagacagagc tggcagacc caatccctt acatccctaa 600  
 caccaggaat ctgacttca agtggacagc ccaagtgtgc agggctccag gagggccag 660  
 gcaaggcttg tatcaggtta ggaagcttg agctgtgat gacatata acatataag 720  
 atacattat aatataca caggtctat atattatat acatgtttc ctggcccaag 780  
 agctcttg gttcagggc cacttcaaaa ccttccctgg ggaggtctgt tcttccag 840  
 gatcccttg caggagga gggagggaa cagggtgtgt ttctccatg aagaggaag 900  
 ctagatcttg gcaagactg catcccatgt tcccatgtc tcttccgtcc 960  
 ccaaggaatg gaaaggcagt tcccttccc cagtggacgt ctagggtggg acaagggtac 1020  
 ttgcttcca gctggaccag agtggccctg ttgctctgc tctcccttg tggggactca 1080  
 ggcagcagag gcatctggga agtctctgag taggcaggtt cctcctggga ggcaacccca 1140  
 cctgtttgaa aggtctggcc agggctgttg gttcaggctt gtaattccag caacttggga 1200  
 ggcaggtgag gggaggtcac ctgaggtcag caagctggc caacatgatg 1260  
 aatgtgtc ttaactgaat atgcagta 1320  
 cccagctaca g

<210> 61

<211> 3227

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_Feature

<223> Inyte ID No: 1752768CB1

<400> 61

tccagtacc tccatgtcc cgttgtctg gacggcggca gtcggcggca gtcggcggc 60  
 aaacagcgat tcttcagat tgcggggaat taaagaagcaa ctgctgctta ttgctgggct 120  
 taccggtagt cgggctctac caaatgtctg gcttccggcc attaagaggt gtcgaaggt 180  
 ccttcagctg ccttcagacc agcttggccc gcttccggcc attaagaggt gtcgaaggt 240  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 300  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 360  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 420  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 480  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 540  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 600  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 660  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 720  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 780  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 840  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 900  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 960  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1020  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1080  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1140  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1200  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1260  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1320  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1380  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1440  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1500  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1560  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1620  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1680  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1740  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1800  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1860  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1920  
 ggtatagatg gcttcagacc gttcagaggt gttcagaggt gttcagaggt gttcagaggt 1980



cttgcagtt aagaatctat ctgtctcat tatgtccaaa gattggctat agatccttgc 2040  
 agacagaga atgcctcagg aacaagttt cccccaagaa agtccctt agtagcattt 2100  
 ctcttccct gaatgcttt tttttttttg cccaagaa ccaatctgtc gaagattgaaa 2160  
 agacagaga ctctcagg aacaagttt agtggcagt agggcaatga gtaaccctag 2220  
 aagtctgcca ctctcagg agtggcagt agtggcagt atttgcttgg atttgcttgg 2280  
 agcaatgagt acaatgacc ttgctcttgg atttgcttgg atttgcttgg atttgcttgg 2340  
 cagagattag gtttaaatgg aatagagctt aatagagctt aatagagctt aatagagctt 2400  
 aactagatt gaataaagaa agggcaaaaa taatctcaca gaagcttggaa gtagcttggaa 2460  
 cccctactga gggccaagaa ttgctatata gttgctatata gttgctatata gttgctatata 2520  
 agacagaga ctctcagg acaatgcttga ttatagagaa ttatagagaa ttatagagaa 2580  
 agaatataac tcaatcagg ctgttcttgg ttatataatt ttatataatt ttatataatt 2640  
 ttttccctcc catctcagg aagaatacat tttttcaagg ttatataatt ttatataatt ttatataatt 2700  
 ggaatcatlc ttatgttggc cttagagagt tttaggcttgg tttaggcttgg tttaggcttgg 2760  
 gtaatatgc ttatatgttc gttgtaatat atttgacagg tttaggcttgg tttaggcttgg 2820  
 gtagctcagg atctcatctt atctcatctt atctcatctt atctcatctt atctcatctt 2880  
 aacatagac atctcatctt atctcatctt atctcatctt atctcatctt atctcatctt 2940  
 catctgttac taagctcttc caggcaagct gtatatagtt tactgttctc tactgttctc 3000  
 cctttcaag gttccatgg ttcaagaatga tgtttgattc tttaatrtttc ttgctcttct 3060  
 ataatrttct tttaattgatt ttgctatatt ggaattcaat aaaaaatttg aacaataata 3120  
 tcttaatat aactgtttt gttgcatag aatcatata agtaataata aaaaaaacaac 3180  
 aacatagagt tactatagtg gttataatat aaaaagttaga aaaaagc 3227

<210> 62  
 <211> 1865  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Inyte ID No: 1887228CB1

<400> 62

gttctagatc gtagagcgcc gctcccgga gttccctcg atgcccctag aagaagcagac 60  
 tcagaaattgg gcccaagagt gaagcccaag aagcccctcc atgcccctcc atgcccctcc 120  
 ctagaattgag gtagcctgctt ctagcctgctt ctagcctgctt ctagcctgctt ctagcctgctt 180  
 aatgattagaa aggttgcctga ggaactcagg gtagcctgctt gtagcctgctt gtagcctgctt 240  
 gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 300  
 aagttcccaag gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 360  
 cccagcagaag cgaactcatc ttgagcccaac cgggccaaga acttggcctat cacccttgcg 420  
 aagggccaac ttggggcgga gctgctgctt gtagcctgctt gtagcctgctt gtagcctgctt 480  
 ctggtgcttgg aacttccctg gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 540  
 ttcatgctat gctcagcgcc catcccgccg gtagcctgctt gtagcctgctt gtagcctgctt 600  
 ctgggccaac ttcccggaac agcccagctg ctcatgcttgc aacttgaaatgc catcatctga 660  
 gcttccatgt ccatcgaagt ctcttgcaaaa ctctgccaag ttcttgagat ttgcttgcg 720  
 tttagcaact acatgaacag tagcaagcgt gggtcagcct atggccttccg gtagcctgctt 780  
 tttagcaact acatgaacag tagcaagcgt gggtcagcct atggccttccg gtagcctgctt 840  
 ctggaattgct gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 900  
 ctggaattgct gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 960  
 ctaggcttag agttgacaca gtaggagttt gttgctgctt atgacctgctat gttgctgctt 1020  
 gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1080  
 gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1140  
 gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1200  
 gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1260  
 gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1320  
 tttaggcttga aacggagctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1380  
 gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1440  
 gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1500  
 ctggaattgct gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1560  
 ctggaattgct gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1620  
 gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1680  
 ctagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1740  
 ctagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1800  
 ctagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt gtagcctgctt 1860



1865

```
<210> 63
<211> 1924
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<223> Incyte ID No: 1988468CB1
```

[illegible]

```
<210> 64
<211> 948
<212> DNA
<213> Homo sapiens

<220>
<221> misc-feature
<223> Incyte ID No: 2049176CB1
```

<>400> 64



aacttggaag gcaaggtta aaccttgcaag ctactgtgc tgttaactga gcaagtaagaa 480  
gccaagagac aataactgaac cgtttaagaa gcttcaglat gcatgtattta acaactatcc 540  
aaagtgtaga gccgtgtgga caaagacacat accaagaaat accaagaaat accaagaaat 600  
gtacacccag cccagtgaac tcaagcaggt atggaatccc accaagaaat accaagaaat 660  
gtacacccag cccagtgaac tcaagcaggt atggaatccc accaagaaat accaagaaat 720  
tccgaagatc gcaaaagcatg aaatcctgtga aaaccccaaa agggccgcaaa gtagtctcga 780  
accggtcact aaaaatacaaa gtgaagaaac gaaccaagat gtaatttag tcaatctcta acatgataa 900  
gtcaaatatc ccaagacaga ttatgtcaaa tcaatctcga acatgataa 948

&lt;210&gt; 65

&lt;211&gt; 2035

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Inyte ID No: 2686765CB1

&lt;400&gt; 65

gaccgtgtgc ctgaacgca aaccccgct tgcacccaag ccggtgaacca cagtgtgcctc 60  
agacagagtg atgcagagaa agtgaagaaat gggacaaagt gggccctcga ggtccctcga 120  
aacatcctga aggtccctcga gtagcctcga gtagcctcga gtagcctcga gtagcctcga 180  
tgggacttgt gagctccggt tagctgtcgt gaccgttgaac gaccgttgaac gaccgttgaac 240  
aaactcctaa gaccgttgaac gaccgttgaac gaccgttgaac gaccgttgaac gaccgttgaac 300  
gacagccccg gacacgctga gttacttga ctcagtgtag ctcagtgtag ctcagtgtag 360  
gtccttgcgt gatgatgtga gctccatggt ctcagatcca gtagatcca gtagatcca 420  
gctgcacagag gacaaagtatg gctcctcctg ggggacagcccc gtagcctcga gtagcctcga 480  
ctcatctccc gtagcagctg gtagcagctg gtagcagctg gtagcagctg gtagcagctg 540  
ctggtatgaag tgcctgtcaac ggcgatcca ggcgatcca ggcgatcca ggcgatcca 600  
ccctcctcct ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 660  
gcaagaaccca ggaagtcttg aagccctcct aagccctcct aagccctcct aagccctcct 720  
tggtatgtag aaggaacctg aacgccccgt cccctcctca cccctcctca cccctcctca 780  
gtagaatacca gtagaagcttg aagccctcct aagccctcct aagccctcct aagccctcct 840  
ccctcctcct ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 900  
gtagaagcttg aaggaacctg aagccctcct aagccctcct aagccctcct aagccctcct 960  
gtagaagcttg aaggaacctg aagccctcct aagccctcct aagccctcct aagccctcct 1020  
tgcaggggctg gtaggccccatc agctgtggaacg gtaggccccatc tctgcacatcc tctgcacatcc 1080  
ctcccccgtg ggcgatccgc aacctgacgag gtaggccccatc gtaggccccatc gtaggccccatc 1140  
ggagtgatctc atgtgtcatctc tccgccccgtg cctgtccctg gtagcctcga gtagcctcga 1200  
ggacagctctc tctgtgtgaag gctgtcaagat cactcctccggt gtagcctcga gtagcctcga 1260  
ccaacagctg ggcctcagtg aagagctcgt gtagcctcga gtagcctcga gtagcctcga 1320  
gtagcctcgt caactgtgaag cactgtgatga gtagcctcga gtagcctcga gtagcctcga 1380  
gtagcctcgt caactgtgaag cactgtgatga gtagcctcga gtagcctcga gtagcctcga 1440  
gtagcctcgt caactgtgaag cactgtgatga gtagcctcga gtagcctcga gtagcctcga 1500  
cactccctgtg ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 1560  
ccctccctgtg ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 1620  
gtagcctcgt ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 1680  
ccctccctgtg ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 1740  
gtagcctcgt ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 1800  
gtagcctcgt ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 1860  
gtagcctcgt ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 1920  
gtagcctcgt ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 1980  
gtagcctcgt ctcagagacca aagccctcct aagccctcct aagccctcct aagccctcct 2035

&lt;210&gt; 66

&lt;211&gt; 766

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Inyte ID No: 3215187CB1









ttactcata ttataaaga ataagaaa atctagata attctcaat ttgatagaac 2340  
 tgttcagcct ttcaagatt tcttatata caaatgatta catataatg aatgtacatt 2400  
 ctctcactg actrtggtga tttagaaac tagaatgattg tgrttctatc tgraatattc 2460  
 ttccattga aaaaatctc aaaaacacaga ttaaaaccac aaa 2503

<210> 68  
 <211> 541  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Inyte ID No: 5080410CB1

<400> 68

atgggttcca tggggagag cgaacagggc ctgtgtctgc acaacagct gggtggccag 60  
 gacgaagctgt gggtggccgc cagcatcggc tccctgtcca cggcggcggc catcgacaac 120  
 atcgtctctc gcttccatgg cctcctgtcg gcagtgtaagc tcaagtctgt actcggggacg 180  
 ctgcaccctc cggcggccag ggtggacagc catccctatt tgccaatgaa gggtggcccta 240  
 atggaatcca tccgtctcgc cagcctccat tgggtgtccat gggtggccgaa 300  
 atcttgaaat ccttccggga cagaagctcg cttaaccctg agcttggaaga gcaagaatccc 360  
 aacgttcagg atattttggg agaacttaga gtaaaagtgtg gtgagtgtaga aggtcttgc 420  
 atgtctggcc atgtctggca gtaacttgaa aaaaaaggccg ctgacggacc tcggcgggacc 480  
 cctcactccc cgggttgaa catcttcagt taaagcggaa acccaagagc gccacgctgc 540  
 541

<210> 69  
 <211> 937  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Inyte ID No: 5218248CB1

<400> 69

gactacgacc aaaaacaagg agcgcggc gggtggcggac ttaccctacc ttctctgcct 60  
 tccggtcgct tctcagccgg gcgcggcggac caaaggagcc gtccgacatb gtcctaacatg 120  
 gagaaaaacc tgttcaacct gaagtctcgg gccaaagaa ttgagttagag tgccaaaaaa 180  
 tgcgaataag aggaagaagg cgaaaaggcc aaaaataaaa aggcattca gaagggcaac 240  
 atgggaagtgc agatgtgcga gtgcctgcga gggtccagac gggtgtgacg 360  
 atgggcaagg tgaaccaagtc gatggtctgt gtgtgttaagt cgaatgtatgc gacattgaa 420  
 accatgaaatc tgggaagat tctgtcttgg atgggcaaat tgcagcaaac gtttgagact 480  
 ctgacgtcc atgacgagca atggaaagc atgttagac gccagcagac gttcaccact 540  
 ccccaagacc aagtgtgat atgtgtccag gaaatggcag atgagcggg cctcgaccct 600  
 aacatggagc tggcggcagg ccagaccggc tccgtgggca cgaagcgttggc ttcggcgggag 660  
 caagatgaac tgtctcagag actgcccgcg ctccgggac aagtgtgacg gcagaaaccg 720  
 ctctgaggt tcttggccat agccaacct tgaatgtctc tctgtgtgtt agagagatac 780  
 tataccctag aaaccttgaa cagccagaa tgcgtgaaat tgcgtgaaatg ccttctacc tttaggtta 840  
 cagcccccac caataatc aagaaatca gtatctcgc acctctagct gatatctaaa 900  
 gttctgtata gctcgtaat atgtatttt tatagca 937

<210> 70  
 <211> 823  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Inyte ID No: 058336CB1

<400> 70



ccatcaagg cgtagtgcga gctaaatca accctcata aagggaataa gcttgccg 60  
 gccggcgga tggctgcgag ctgcgagcc gctaaatca accctcata aagggaataa gcttgccg 120  
 tccggctgag aagaagaaac cccgcttga gagtgaagatg gcgttcaatg attgcttcag 180  
 tttagaatac cctggcaacc cctggctgag gaaagtgtcc gtaagtgtcc 240  
 ttagcaactg gacctgtatc tgggtgatgg ttagcttatc aacatagcac ctgtatagatg 300  
 catctctgcg tcccttaca gcttacaag tgtatctagc agtaagcccc tgggtgaataa 360  
 gacatctctg aaggatgttg tgggaataa catatacaga aataaacaata aatacagatg 420  
 aacgtacccc cctcctccg tggaaagaa cataaagcgg tcaagtttg caattgaca 480  
 ggaagtggcc tataactac tgtcaaca ctgtgaacat ttgttgacat tgccttcgca 540  
 tggagaagg gtttcagagc aggccaaacc agcgaataag accgtttagt ttgttgacagc 600  
 tgcctgttgt gctctctcat tcttgagc gtctccaag ggaagagag atccctatga tggatggcag 780  
 ggggtggaata ctatttca gttgcatcat acgttccag atccctatga tggatggcag 823

<210> 71

<211> 1033

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1511488CBI

<400> 71

gctcagctgc tccagctgc acgtccagc cctccagc cctccagc cctccagc 60  
 tgcaggagac tgcaggagac acgctgtg cttggttcagg aggttcaggc gcggctggga 120  
 acaagcggc agcatgtca gggcggctcagg gtagcctacg aggttcaggc gcggctggga 180  
 agtccgcctgc gggcggctcagg cggcctccga gggcggctcagg agcaatctcg gggcggctcagg 240  
 gcccggggt ctcctcgtc actccagcgg cggcggctcagg gggcggctcagg agcaatctcg gggcggctcagg 300  
 tcccaaggc tcccaaggc gggcggctcagg gggcggctcagg gggcggctcagg agcaatctcg gggcggctcagg 360  
 cctgcctcgt tcaaatctcg tcaaatctcg gggcggctcagg gggcggctcagg tgaatggac 420  
 aatgatagac accggaagac acaagctcg agtgaaagct tgttacaataa tgaatggac 480  
 caaatatc gctgccttg ctgtgatag tgcagccaag agctaaagt gttcaacctga ctgccttg 540  
 atccttaaca agttggaact tggccaagaa agctaaagt gttcaacctga ctgccttg 600  
 tgcacagct aagctaaat gatattcaa gttgaacaagt gttcaacctga ataccatccc 660  
 tgtcatcagc aacagtgaat gatgggaata atagatatc ttttcttcta taccataataa tctgcctag 720  
 ttttatcttg gtaacaaggaa gcaacaatgct ttttcttcta ttttcttcta gtaaaccttt 780  
 actgaagat accatgcat caaaaagtg acaaacctgt atacagttctg atagatatc 840  
 atgtcgtgaa cactgtgtat accactgcca aagtgaagat gtagaataat ggcacaactt 900  
 cccagctcca tctcctgccc tctcagccc tctcagccc aacatagcag ttttcttag 960  
 tttcatcac tttgatcat tttgctgtt tctcagccc taccctccca aacatagcag atttatcat 1020  
 taaaaaaaaaaa aaaa 1033

<210> 72

<211> 1622

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1638819CBI

<400> 72

ggcacttccg gcgcgcgct gcaaggcgg gaaacaccaa tggcggggtat cttagaagctg 60  
 gtgtgtgtt cctttcagcg tcaagggtc cacactgttg gtaggtcgtg caagaaatcg 120  
 acaaggcctg agcaacctgt gctgaacccc catctcaggg acccatctgt gaaagcctcg 180  
 aaggttgagaa gttacccgtg tccaagcggc ttcgaagcgg ttcgaagcgg acccatctgt gaaagcctcg 240  
 cagatctcgc ggcggcgct tgggtgtta gactgtgggg cagctcctgg cagctcctgg ggccttgagt 300  
 caggttggcgg tgcagaaggt caacggcggca ggcggcggca cccctgggaag gtagcaactt tctgtgccc 360  
 gtgtctgggg tagatctct tcaatatc cccctgggaag gtagcaactt tctgtgccc 420  
 gctgaagctga ctgaacccgag aacctcacaag aagaatctcg aggtgtcctc tggcaggaga 480  
 gcaagatgta tctcggcgct catggcgagg ggttccggga cctcgatcat 540



gacagagctca tcaagcctgtg cctgaacctt ctcagcgtga cccagacat cctgcaacct 600  
gagcctccctga agagttcttg aggtttcttg gctgtatg gcttttaac gctgtatg 900  
ggagttctct cctctctct cctctctct cgtctatg tgaatgagat gaacaacta 960  
agttcagggg ccatggaaaa tgaataagtc cgtctatg tgaatgagat gaacaacta 1020  
atcaagagaa agaggtgagag atggaaagat agacgtgtgg aagatcaga 1080  
aggaatccgc cgaagcaggg atgggtgtgc cctgtgtgg cttcatctta 1140  
tagactgtta aactgttaca cacaaacagg cttccacc cgtctctgag agcacccagc 1200  
acagatctcc agttcttagt gtggtgtgtt aaagttagaa atctgtgggc tgggtgaggc 1260  
caactcagcc tgttaaaccc aggtcttaga aggtctgagag tgggtgagatg cttgaaagtc 1320  
ggagttcagg accaacctgg gcaacatagc aacacccccc atgtcttaca aatgaaaaa 1380  
ccaagaagca aaccaaaaga aaaaactgaa atttccatc tgggtatcaa tctgtcttt 1440  
ctgttgaaac atatagcaat tcaagcattc tccaagcagc aaaaagtccc ggaacaatta 1500  
gggaagacgt atgtctgaa tttatccagg cagtgtgtct gcttgtgtt tgcctgaaa 1560  
tttatccag tgtctgggct cccaagaaca taaatgtaat tgcacaagca aaaaaaaa 1620  
aa

<210> 73

<211> 2449

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Inyte ID No: 1655123CB1

<400> 73

cgctgcggag cttcccgaa ggaagcgtgg cgcgcctcc cctccacc cctccacc 60  
ttatagttcc gccgcctcc cctccacc gagaagcgg cgcgcctcc cctccacc 120  
ggaagcggag tgcgtgtgtg gagaagcgg cgcgcctcc cctccacc cctccacc 180  
ggaagcggag tgcgtgtgtg gagaagcgg cgcgcctcc cctccacc cctccacc 240  
ggaagcggag tgcgtgtgtg gagaagcgg cgcgcctcc cctccacc cctccacc 300  
tccggcggcg gccgagccct tccctcaact agcgtatgcaa tcccttccg agttcaactgg 360  
ccggcaccg agttctctc gcccaacct ggagtttat ataagaaaga taattatgtc 420  
atcatgacaa ctgcaacata agaaaaatat aaatgcatat tcccccctgt gacaagtggg 480  
gatgagaaag taaagaaaga ttaataagag cctaaatccc gagagctttt gaggccacta 540  
tttaaacaaa gcaagtgttc ctacagaatt gactcttat ggaactcaga agtatgtcat 600  
ggaataacaa tctggcagta ccatgaaagag aaagaactg gtcagaaaat aatatattac 660  
gagttacata tctggaaat gttggccaaag aacctcttat ttgaaaaaga acgagaaagca 720  
taatactctg tgggaaatgg aaatgtgtac cctgttagt ttgaaacagaa ccggcccaaga 780  
taatactctg atgtcatctt gaatcttaagc atttgacac cactctgtg cagtcatctt 900  
aaaatatagtt ccaagcctcc tccgtgtgaa tccgtgtgaa gacatatatt gacaggaatt 960  
ccatttaagc cctccacc gaggcagcgt gagaagcagg aaagaaatat aagggtgtgc 1020  
tttagagaa ataaagagga agatttgc aa tcaactaa gaaagagat tccagcagatc 1080  
cacaaagtcga ttgcttcaact gttgggcaac ccaatatacc ccaatatacc 1140  
agaacagtgaa aaaaactcct caatctcaag ttcgtatctg aggtatatac tgaacaatac 1200  
gaaatgaggg aaagtgaaga tcatgtatct cttccttcct ctcccttcct tccccaacta 1260  
aaatcttag agttcagcag aagaaaaaga tcatgtaaag cttccttcct tccccaacta 1320  
aaatcttag agttcagcag aagaaaaaga tcatgtatct cttccttcct ctcccttcct 1380  
ctctgtgaaat acatatgtgt gaaatgagat gctgtatg gcttttaac cgtctatg 1440  
agttcagggg ccatggaaaa tgaataagtc cgtctatg tgaatgagat gaacaacta 1500  
gagcctccctga agagttcttg aggtttcttg gctgtatg gcttttaac gctgtatg 1560  
ctgttgaaac atatagcaat tcaagcattc tccaagcagc aaaaagtccc ggaacaatta 1620  
ctgttgaaac atatagcaat tcaagcattc tccaagcagc aaaaagtccc ggaacaatta 1680  
ctgttgaaac atatagcaat tcaagcattc tccaagcagc aaaaagtccc ggaacaatta 1740  
ctgttgaaac atatagcaat tcaagcattc tccaagcagc aaaaagtccc ggaacaatta 1800  
ctgttgaaac atatagcaat tcaagcattc tccaagcagc aaaaagtccc ggaacaatta 1860  
ctgttgaaac atatagcaat tcaagcattc tccaagcagc aaaaagtccc ggaacaatta 1920  
aatgtctgtg cctatggaaa gttcttaaat aatagattca ttatgttaat tcatgtatat 1980





gtaagtagct aatgaagtaa agatcatgaa gaaagaaatc gataagtgta aatgagagac 2040  
 catgtaaatc tagtaacctg aatcccttca acaagatttt atatatgcaac 2100  
 tgcctcctgc aagttaataa actagaaact gggcacatgg tagaggtcca catgggagtt 2160  
 gtcctcaacc ttgttaactc caagaaactc ttattataa taggttgctt ctctctcaga 2220  
 actttatct atacttttt tctcttatg agtatgttta ctctcagagt atctatctga 2280  
 tgtagacagt tgtgtatgct tctgagactc agaatggttt actctaaaca aacacctgtgc 2340  
 tgtctatcc ttgtacttgc ctactgtaat atgttatcca ctctggaaca gtttaccaga 2400  
 caatatctat tttaagtgta ataaatgttc cacaagcaaa aaaaaaaa 2449

<210> 74  
 <211> 1689  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2553926CB1

aagtaacttg aggatctgtg ggaagcgacg tgaactcggc gcttggaag atgagagcag 60  
 cggagacaga ggcggagagt gcagccctag aggtccctgg tgaagtgga ggcattctgg 120  
 aacctgttag cctgcagagag gaggcagaa gataccctga gcttgagcc ccttgacagg 180  
 agaacatctc ggtctcagag gacactgtca aggtctcga ccccttgctt tctgaaaga 300  
 agagccgaat gcaagctaaag aacaatgaa agagctgaag gccaacctca 360  
 cgaagcccatc aagagctcatc aaaaatggcc tcccaactca g ccttgacagg 420  
 aagaaacaaa gctcagggcc gctcaaggcc aagaaacaaa ggaagaagcatc 480  
 tggcccatga gaaacgcaga gcaagctccaga aaccttgaa aaccttgagg 540  
 tgcagcatct ggcggagagtt tctgcaagagg tgaagggagcg taagacaggg aacctagcag 600  
 agcttgacgg ggtgttctca aaccttgaa aaccttgaa aaccttgagg 660  
 acaagctgca gaggatcag aacctctcc agctctgta taacctgcaag ggtaaagctgt 720  
 tgtccctcga gactgagagc gaggcagaga atctctcaga tgaataaacc tgaataaacc 780  
 ctgcacccca ggcagagagt acagagaca ccatggggag agaccctggt ggtgcctca 840  
 agttctccaa ggtctgttgg ctacaacctg ctggagatgt aaatctgcca tgacctctg 900  
 gaggacaga gcatggagaa agatccctga aagggccctc gactccctc aacctcccaac 960  
 catcatca gaaagactg tgaacctctg agttcagctt gatctctgac taacatccca 1020  
 caagctctgg catctgtgga ttaaaatccc tggatctctc tcaagtctgt atttgtcat 1080  
 cctcatatgc tggcaggaac aactatcat acagatctc agaaagccaa aacatggacag 1140  
 gagctgggac tggtttgaa acaggtgtg cagatgggg aggggtactg gacctgggccc 1200  
 tccatagtag cagaatctgt gaattaat caagggagag gagaatgttt taggcaggtg 1260  
 gttatagtg gaaagataat ttatctcag gatccaaa tttgttgag tttgttgag 1320  
 tgcataagtt ctgcgtgga accagatga taacagtgag ctcatctgac tgttttagga 1380  
 tgaacagctc agtgttaaca tctctgtat ctttttggc ctatctaaa acaattctcg 1440  
 atcacatggt tcaagatgttc attatata tctctttcaa agatctcaga attggtctt 1500  
 gtcactcact atgttatgt ttgttcat gacctcagt gatacctga tcttccca 1560  
 ttctgttctt cggatvggag aagatctcga actcttact ggtataaataa 1620  
 atcaactcac gttattggat ctctattgt tttctcaaat tttcttga aatatattaa 1680  
 aaaaaaaa

<210> 75  
 <211> 2489  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2800717CB1

tgtgcccaga acgcggttag gaagtgtgtg catagctcg aaccttaat gttctcagt 60  
 tctgtaact tctctccca ctgggtgga tagggctt aagagcagt ggaatgcat 120  
 tccccgatc aggtaccag ttgttgcctg tctgaaactc tgcagctct ggagactgt 180  
 gccctgagct ccaaccagcg ggcctcatcc tacacccctc ccaaccgaac tctcaccg 240



agcaagaagc agctccacga gagaagaac gttccacat gcttagccat gggagagagc 300  
gctgcacagc cggaaagat ccagcaccct gggtcctgaca tgcggcagga aagccctcg 360  
agccccaagc cgaatgcttc ctcacacaca agccccaagc tgaaccctag ggaaccaag 420  
gaggccatcc gggaacatcc acaggtgaaac gcaagtgcagc gcaatgcagc 480  
agttctggagg gctccgctgaa ggacatccca aatgacccca aatgacccca 540  
gctccacagc gcaacacagc gagcagaagc aggtgaaagc aggtgaaagc gctggaagc 600  
aacccagc agctccctcc agccaacagc catgagtagc cagtgcagc aggtgaaagc 660  
aacccagc agctccctcc agccaacagc catgagtagc cagtgcagc aggtgaaagc 720  
gagatcccg agctccctcc agccaacagc gctggaagc gctggaagc gctggaagc 780  
gagatcccg agctccctcc agccaacagc gctggaagc gctggaagc gctggaagc 840  
gagatcccg agctccctcc agccaacagc gctggaagc gctggaagc gctggaagc 900  
gatgatgat tgcacacaga tggagagagc ctggaagaca gtcgagagga aaggtgga 960  
gaaagttagg cagaaagatc aaaaagatcc agcctggaaag aagtggatag cctcaagaa 1020  
gcatcttc cagaaacat cgaagaaag atgaacagc tgggacaa gatcgatc 1080  
gtagagagga gtagaagatc taagaatct ctacagtc aaatccca aatattcc 1140  
ggaaaaagc cccctccaa gttctccg ggcgggaag agtccgagag 1200  
ggagaagcc atgcagaaat tgaagcagc tgcctagcag tgcctagcag 1260  
ccaatgac ccaatgagc gtccttgca gaggctcat ccgaagcgtc cctgcagc 1320  
gctgtggtg aaggtggaat tgcagagagc gctgctgga aggtggaagc aggtggaagc 1380  
aactcggga catcgactg actatgga agatggaag ggaagagc 1440  
gtgcccctg aacagcgaca gaagttagc ttaggtgta gctacgcgt aacatccgag 1500  
ggggcggag gctccgtagt ggaccccggt tgcctcagc cctgcagc 1680  
agaaacatc ctgcacatc tccagcagc ctacccat cctgcgtc tgcagcagc 1680  
gtatcatc agccaatc catagtagc cacaagagc caggtcttcc catgaaagc 1740  
acctgcacg gctcccatc ttaattgct ctgaagtc atctctgtag atctctcc 1800  
agattgaa agaaactgag cagtgaaaa atgctaact cttgacta gtcagaaaa 1860  
aacagagat aattagata ctagtcatga aagtgatc atctcttct gtcattccat 1920  
aagcttgcg aatagtgtag cgttatatcc accgtatct gttgaatccaa 1980  
ttatatct ttaagtgtg atatatata tacaataa tgaagctaa acatatact 2040  
atatgttta agaagaaac atccagaaa gtaaaaaa gtagatcagc tgggtgta 2100  
ctgtcgtaa cccaatgct tatgcaaac gaaagaaatc tgaagagatc aaaaagatc 2160  
gctatgaa gtaagtagc agtaagtc aagatc aaaaagtaga aagaatatt 2220  
gtaaaatgag agaaacagga aacaaagat gccgaagaa atgaaaaatg 2280  
tatgtctg aagtaaat atctagat atctagat gtagcttaac tatctcactc 2340  
aatactaa agttcgtat atcttgat accctgagat aacatgcat cttcaaat 2400  
ctaattct taataata catcttag catcttag tgttaagc gtttaagc 2460  
aggaatctac aatatctc ttaaaaaa

<210> 76  
<211> 898  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Inyte ID No: 5664154CB1

<400> 76

ctctggcagc tgaacaagc cacaagtagc cgcacatgac cgcacatgac gctccctc cgcacatgac gctccctc cgcacatgac 60  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 120  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 180  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 240  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 300  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 360  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 420  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 480  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 540  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 600  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 660  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 720  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 780  
aggaagagc ggcagagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc gtaggaagc 840



gctcattata tattgtgaag ttaataaaac agttttaaaa agcaaaaaaa aaaaaaaa 898

<210> 77

<211> 1236

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 017900CB1

<400> 77

cctcggatca gaactctga gagccgggtg gagccattg acgtccagcg acgtccagcg  
gacgcatgga aggtctgggt gacgtctgga aaggtccctc tggccttgcc gatccggcct 120  
cgacacggctc gctggaggaac ccgatctcc tttcccgaga cgttgagcg cgtataccgac 180  
cgactcccg agttcatcgt gcagacgggc tccatccggc atcacccggc tccagggggc 300  
tgggtgatcc cctacatcaa gaaggagagc tggattggag actcttaggc cgggaggaacc 420  
gcccagatga agcgagtcct tggatggggag gttccgctgtg ttacctggcg ccgctcgggt 480  
ccgcaaccggc gccctccctc ccgctccctc tccctccctc gcccgcggga ttctcccgct 540  
gctccctgtc cctcccgctc agtgcctgac ttgtttccag gaattagcgct ccaggtcctc 600  
gctgcccggc cttgggacctc actgggagcg agccgcccgc cttccctcc agccagggcag 660  
ccctcccat gtacatttgg accgtgtcct ggcgtccagc ttgcaagctgg gttcccttta 720  
cacacttggac agacacaccaa ctgcccggc ttgccaaaggcc tctccctccc accagactgc 780  
cagacagatca catcatctctg ccacagagac ttgcgtctggcc cagccatccgc catcccatcg 840  
atccacagca cagactctgt cctcatgtga tctggactca cttggaggt atctgggctg 900  
gccacagtcc ctggacagtg atccagacag ctggccggcc cccaagggat ctgtcacct 960  
cagcgaggaac tatctccctc cacccccag aaacctcttg ttgtcttgcc tagggcccag 1020  
tgttccctgc agccaatcg agtctctcat ttctctctgt ggaaccagtt gatcttgcca 1080  
tcaacacagt tgcctttgt tgcctttgt tgcctttgt tgcctttgt ccttatttct 1200  
aataaatgtc agacattatc gaaaagaaaa aaaaaa 1236

<400> 78

<210> 78

<211> 1634

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 035102CB1

gtttgactc cgttgcgtgc ggccagcag ccaaaagct ccgcttgcca ttgctcctg 60  
tactcccgcc gtacactggcg ctgtccaacc cttcccccgg ggtcttgccg gcgtctcgcc 120  
caacccctcg cccgtgtacg cgtctctgac ctgctctgccc ggtcttgccg gcgtctcgcc 180  
agagagcag gggtcccgcc tcccgccagg cgaaggggag aaggatggcg gccatgtatg 240  
tgaagcccgg gctcgagagc cctcgacag gcgcccacag gaggaagctg gcgcccggag 300  
ggaaggagg gctcgaggag ccgcccctca ccggaaagcg gcgaaagcg cgtcgagggc 360  
atcctcgca agtcgaggg cctacccggc gctgaatcg accgttgccg acgttggtgc 420  
gtcctcctg gtgaaagaca agaagaagag tcccatcaca cgtctggaga agtcggtgga 480  
cgttatgga gacttggaaga tctgttccc ggacatctcc gcaaggcgcc cagagcattc 540  
gcgttatgct ttgtgttttg agcttgaaaa gtttgaccgc atctgacccc aagcaaccac 600  
gatcaacaac ctaaaaacct ttgagagag gagagagag gaggtatctg gatagatgg 660  
ccccagattg gttctgttaa ttgatgtacct gggtcttacc tatatgagag gtataagcgc 720  
caggggaggc caggctctgg agatgcctgc tccggttgggg gttgcacaacct caaagtata 780  
ttctcctctt gggtatccga agaggtctat tatgggaagt ttgttgagcg agcgtatctc 840  
cagttacagc cgggtgtccctc aacccaatcc gaattcctct gggtccccc 900  
aaggcaacct ggaaatccaga agatggaaat gttgggccaa gttggccaaa ttgcataaga 960  
cagaacccag cacttgccag ttgagtaacc ccttgggggt ttgagtaacc ggtccgggcat 1020  
cagagccaag gccagagctg aagccaagct gaggggccag ggcagtgcta ggccgggcat 1080  
ccaccctctg tgaagggttg tgaagggttg tgaagggttg tgaagggttg acgaaactact 1140



gtcctgagtc atagtaata tgggtggggc gagggtctta ttctgtaga aatcgtgtga 1200  
 cttaaggat ctgattttg taccattatg ttgttaacat ttaataata ctgttaaat 1260  
 gctgtttgt aatgagattg gctacattt tctgttagga tttaattga gagttrttgtc 1320  
 gttttgtta aatgattga agaatttgt atttattg ttgtattga cagattatgc 1380  
 aacatttgaa ggaaggtgt actttgtg ttgttagaa cttagcagta tgaattatgc 1440  
 gtctcagggg aaaaaatag ctggtttgtg tctagccccc caaacattt gttgtttgtg 1500  
 tataaagaa gaaagattg catgtacct catgtctta gctatttga tatcttagga 1560  
 aaaaatataa tgaattagt tagcagttat ccttgacaa cttaataaat taaacattg 1620  
 tggaaaaaaa aaaa

<210> 79  
 <211> 1258  
 <212> DNA

<213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Inyte ID No: 259983CB1

<400> 79

tccggcag cgttagcaac tgcagaactg caggagacta tcttctaga caaggcagt 60  
 gaggaggagg gaggcctga ggggtgactg cctggcgtg actccgcac tgggggacat 120  
 tatgtgcgt ggaacgctg ctttgggaag gcaaacctc ctgaaatggac catgtacctc 180  
 accaaagatc cctgtctctg atccaacaa cagcttcaac cctgaaacaa ggaacgagaag 240  
 accaaacat ctgaattgga agcttaattga cctaaattga cctaaagact cagaacaggc ctactatgc 300  
 ccgaagaag agtatgttgg ttctcaaga ttctcaactg accgaagtat accgaggtgc agagggtgc 360  
 ctgaattga aactgcagaa aactgcagaa aactgcagaa ggaatttaaa ggaatttaaa 660  
 actctatc ggaacaggat aaaaagcaac cagtatccac ttgaactca atctccctg 720  
 taccgttaac agttagatga cgttagctca cagtaccac ttcaagtgcct cccagctca atctccctg 780  
 ccagttttt ccttttaga tctcaacta ttgaaaaagac agaaatatg ttgtgggtatc 840  
 atctataag gctgtttgg ggaagtctc atctcaac atctcaac atctcaac gcttgcctg 900  
 agcaataaga aagcagctgc tgaagaacca gaggagcagg ggcagagcc tctgcccac 960  
 tccaacagg agtggtagt gaggtttta ttgtagaagg gaaacaaaaa aaaaatatc 1020  
 gaattgtta aacacaaaa gctacaact gaccccttt ttttttttag agggagttc 1080  
 gctctgtta ccaagctgg agttagctg cgttagctg gctcaactga actccgtct 1140  
 cccgggtgtc aggtgattc atagtgccc 1200  
 gtaacagac cggctaatt tttagttta gtagagcagg ggttccacca cgtlgggc 1258

<210> 80  
 <211> 2223  
 <212> DNA

<213> Homo sapiens

<220>  
 <221> misc-feature  
 <223> Inyte ID No: 926810CB1

<400> 80

aaaaagccgc gcggtgctt taggaacggc gctgcctcgt cctgcctcgt cctgcctac 60  
 cggccttgcg aagcagctc ttcgggcagc cccggtctgc ttacggtcca aggaagctc 120  
 agttcttgc cgcctgcaag gcggaagaa gaaaggcggaa tccaacagctg gcaacgctg 180  
 agcatctgct gtcacaacgg ggaagcaagg ccaatcaagg cgtgttgaca aggttgacct 240  
 taccgccggg aggaagcttc tcaacggagg cactgtgtgc agaggtctga aggtgaaataa 360  
 agaacgcgct ttgtttcaga gttcgtcccc ttgctgagta ggaagtcagg gccaccctc 420  
 ctctctccc accgtcagat taagctttc taaaaagct aggtcattc ttatatcag 480  
 ataacctac gttgtcagtc atggttagca tcatgtcagc ttgtcgtaac agccggcggc 540  
 tgaatgcaac cttgcccgtc ttgcccgtc ggcagcagc agaaagtctcg 600  
 gtccataat ggcagcagc ggcagcagc ggcagcagc aggggtggtgc 660





```
<210> 81
<211> 1370
<212> DNA
<213> Homo sapiens
```

[illegible]



<211> 1541  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Inyte ID No: 1496820CB1

<400> 82  
gtgttaagct gacaatact gtacagaata ttaattttt ctttctatt ctttctatt  
60  
aagatgtgtg ttttttcat agcaacatga accgttgaag ccggaattgt ctgcgtatga  
120  
aagaaacggga aagcgcggaat caggaaatct agcaggggcga agacgccttc ccaacttagt  
180  
ctctctctt tgcagagcca taacaagtta cttagcgaagga agatgaagta agataagtta  
240  
ttcaggagttt gcttggaaac ttagtggaatt ctagggattt ctagggagagc agatctatac  
300  
ttcaggagttt gcttggaaac ttagtggaatt ctagggattt ctagggagagc agatctatac  
360  
caaatcttct tgaacagaga catggaggtc ctcatcagag tagcaaatgg actccagtag  
420  
gaacccgcac cagcaccttc cagtcctaga aacgggtctc aggcctacag agtggtacata  
480  
gtagccagcg gacagcgca gttagcagta gtgycactaa cagttagtgt cagagggcag  
540  
accgttagtc atataacaat agtggagca gttagccgga aaaaaggccag catggtatcag  
600  
aacacatcca atcacgttct tccagccctg gaaaaacccca ggctgtttct tcaataaact  
660  
ctagttcatc cagttctcat gggaatgac accatagcaa ggctgtttct tcaataaact  
720  
cacttcggga ccttgatgca aactgggat ctcttcctg tgtacccttt tcaagtgggc  
780  
agcacctcaac tcaatcttct ccaacctcat tgaatgtcaaa gtccaaatca atgtgtctc  
840  
aacccactgc ctatgtgcgg ccatgtggag gacagggatc catggaaaca aagctgtctc  
900  
cttaggcacta cagcagccaa tccatggca acagcatgac tgaagctgaag cccagcagca  
960  
aagtgactct caaccaagctg aaaaatacct cccaacacat ggatgacatca gcttctgtg  
1020  
atgttagctg tgttagatga atccataaag agatgacgca tcatgtgcct cccctctaa  
1080  
cgcctatcca tacaactatc aaaaatgtaa acttctcca tcttctcca actaagttaa  
1140  
gtaaataaaa tgaatcttcc ataatgtlaag aaaaatcctaa atggcttgaac taataatcata  
1200  
tgaatataaa atgtcttgc catctctat ctagtgtggag acagacagta aataagttga  
1260  
taaatagata aatcagata gtgaacaatg ttagtaagat aatagcaggt gtaattggaac  
1320  
tgaagagcatc ttagatcaag aggtatlaag aagcttctga aggcataatg cgaagaggaat  
1380  
tlaaaagaca ttaaacacgc cggacacgtg ggtctaccc tgaatccca gcaatcttga  
1440  
agcttagacc aagtagacct tcaagcatggc tcaagcatggc 1541

<210> 83  
<211> 1372  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Inyte ID No: 1514559CB1

<400> 83  
cggctcagc agctgcgga gtcatltctc tgttgagccg gacgtgagcc gagctgagcc  
60  
cggatctcgc cggatctcagc agcatctcagc tgcctcagaa ccgtctggagc gacgtgagcc  
120  
accatctgcg cagcaagggc tgcctcggcc agtgcagacc agtgcagacc gagcgtcaga  
180  
gagcgtcaga gacgtgagcc agtgcagacc agtgcagacc agtgcagacc gagcgtcaga  
240  
gagcgtcaga gacgtgagcc agtgcagacc agtgcagacc agtgcagacc gagcgtcaga  
300  
tgcacatctg tgcacatctg tgcacatctg tgcacatctg tgcacatctg tgcacatctg  
360  
gtgctccttc ccaagctctac ctccacatc gtgtcgtctg gtgtcgtctg gtgtcgtctg  
420  
gtgctccttc ccaagctctac ctccacatc gtgtcgtctg gtgtcgtctg gtgtcgtctg  
480  
cagaaatctga tcttcttca tttgagaagt aaaaagatga cacttctgaag gggtctcacc  
540  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
600  
gacccctgaag tgaacacagc ctaggctga gctggggacc tggtaaccctc ctggctctg  
660  
ataccccctc tggccccctc gacctgcact aagcaggggg gaaagtgtggg tccctggagca  
720  
cctgccccctc tggccccctc gacctgcact aagcaggggg gaaagtgtggg tccctggagca  
780  
cctctctgac tccctctga ccttgaagga gcttgaggacc tggtaaccctc ctggctctg  
840  
tctgcaaatc ccaatcagctg tcaacagcgg tcatgaatct caccctcttc cctggacact  
900  
cagaaatctga tcttcttca tttgagaagt aaaaagatga cacttctgaag gggtctcacc  
960  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1020  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1080  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1140  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1200  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1260  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1320  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1380  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1440  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1500  
gagtggtggc atcatcaaaa acttggagt cccctcacc cctctaaagt tgggcagaggt  
1541



ggggttactc ctgtgttagg ggtatagat ggggtagtag atctcttag atcctcttag gaggtagaca 1080  
 ctggccctc aatcgtcca ggcacctcc tcatccacc catccctcc cagtccattg 1140  
 caattgat agcagcgga gacatttaa gatgttgca gtagagcta 1200  
 tggacagggc atgcccagtg ggtctatag gggcctggag tagtctgtct tccctggacc 1260  
 aacgttagc cctggaggc actgaagtc tagtgtact tggagtattg gggcttgacc 1320  
 ccaaacact tccagctct gtaacatact ggcctggact gtttctctc gg 1372

<210> 84  
 <211> 868  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_Feature  
 <223> Incyte ID No: 1620092CB1

<400> 84  
 catgaagtc accagcgcca cgtccccgg cgaagtctg cgtgtctgg cgtgtctgg cggagttagc 60  
 agtggccatg gggagcctca gcggtcttgg cctggcagca ggttggttca tgtgatcctg 120  
 gttaatggaa catatgtggg atttatgggg tgcacagggag agagatcagg cttgacctga 180  
 gtagcagctg gtagagctgt ttaggttatg tgaagtac tttaagagc 240  
 tgttggcag tgcattatg agaaagctgt ttaggttatg tgaagagat gttccctcat 300  
 ctctaggct taccagagc tctgatttga agagaaataa tggatttgc acaaaaccac 360  
 aggaagtcc cggagctca tcccgcaat accacagagt gcccttacc gaccttacc 420  
 atggcagaa aagatcttc atattgttcag gtcgcttcaa aagggagat gaaatccca 480  
 agactgtctc gttggagatg ctgtgagctg caaaggaaaca gattatggag ggcacaggc 540  
 atctaagc tgccttagc gttgttagg gcatctcat ggttatggag ggcacaggc 600  
 cgtcccaag acacgagat ttaacaaagt tgaacttaga aagaaagct cgtctgaag 660  
 aggaagcgc tatgaagcc aaaaacagag agcagaggtg acagaggtg tccgtgttg 720  
 aaatccagg aattatgta taacgttgc gttatataa ggtatgtgta tgaagatcca 780  
 ttctataag tatgatttg ccaaacctgt accattccg tattctctgc gtagaagtag 840  
 aataaattc tcttaataa aaaaaaaaa 868

<210> 85  
 <211> 3388  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_Feature  
 <223> Incyte ID No: 1678765CB1

<400> 85  
 aactgtgtc gcaagcagat tgtactcagt tgtcaagag agcagcttcc agcagcttcc tgtttgat 60  
 atatacttc atctctgaaa aagacatgc tagttaaat tcaagatgt tcaatagaa 120  
 gatgacaagg aagatgaggt ggttgtaaa ggtctataa ggtcattaa gtaaaactca agatgttac 180  
 catgacaagt cccctcttgg tatctttrct caaacatga attatgttgg acagcttggc 240  
 gggcagtgta ttgtcaactg gaaggaactc tacaagggtc ttaaccaggc caccctctc 300  
 ggttgcattg atgtcatctg ggtacagcag caggtatgca gttatcagtg ttcaacctt 360  
 cactgtcgt ttggaagct tggagttcc agatccaag agaaagtgt tgaatataga 420  
 cacaagtgga gttgcacatg tcttcaatg aagttgggtg ataacgggag atcttggcac ctcaacca 480  
 gttcagagga atcgaagaa atatgaaaag ctctctgtc atcttggcac agcagcttcc 540  
 cctactgaag atcagttctc taaagatat gacacccctc ttgtgaaatc ggttggagat 600  
 gaaacacat ctcaagttc agacatctca cactgtctgg aaacagagac aattttact 660  
 ccaagttctg tgaataaaga aaaaacgaag agaaagaa atccttga atgttagg 720  
 gaaagcagg ccgcatctgc tgtgcagaa gacacatgtg atgttaggct gagctccgat 780  
 gatgacaagg gggcccagg agccaaggga cttcccttga cttcccttga agaaagaa 840  
 ttgtaagagc cttgtctctc ccatctctgg gatcatcc gttatctga agaaagaa 900  
 tccctttag agaacacata tcccagaca gcgtgtccca agagtgtatc agagcttggag 960  
 gttgaaacctg cggagagctc gctcagatca ggtatccca ttggagtggag gttgggaggg 1020  
 ttcccagat ccaaccaagt cagcaaaaag gaacgatctg accatctac ccagtggag 1140  
 acaattacac catcagaaa tactcatctc cgtgtatga gaaagccaaa 1200



cagaagccc tgggtacaca gatgaagcagc caaacatctg tggcagagct tctcgaaact 1260  
 cctccttgaga gtactcagat tctactctatg ttagatctg accaccttcc caacgcagcc 1320  
 ctacgcggagag cgccttcaga atcccaaccg gacgcctaaag tagactctgcc gtccaagaag 1380  
 aaagtgcttc acaaaagaaat ccaacacccag ggaaccttgatg atatcttaact tgaatgaacta 1440  
 aagggtccctg aacctgaagt tgcagctctt tatctcccta aaagtgaatc agtccccaca 1500  
 tccaggcagtc ggcctcgaagc tgaacacac cctcgctccc agtccccaca gtccgttggga 1560  
 agcgcagctg cagatagctg cctgcacatg cctgcacatg cctgcacatg cctgcacatg 1620  
 atgaaatctc agaaatctc agaaatctc agaaatctc agaaatctc agaaatctc 1680  
 atgaaatctc agaaatctc agaaatctc agaaatctc agaaatctc agaaatctc 1740  
 atgaaatctc agaaatctc agaaatctc agaaatctc agaaatctc agaaatctc 1800  
 cttagcttgc aagtattcca gaagagcttg cctaagcca cagttagtc ctggttgaaa 1860  
 gaagaatctc caaagaatc tggctcgttg tggctcgttg tggctcgttg tggctcgttg 1920  
 aaacagctgc cagaatccaa ggaagggaaa tcttgagagc tcttgagagc tcttgagagc 1980  
 tccagctcca agagcccgcg cgttgccagc cgttgccagc cgttgccagc cgttgccagc 2040  
 ggaatcaga agctcgaaga atccatcaca gtggaaccca tcccacaga gcccctgagc 2100  
 caccgagcga caactcaca taagaagctc ctcgcctcc ctcgcctcc ctcgcctcc 2160  
 ctgaagctcc aagatgagtc aaatgagtc gtgtttagta ttaacaaacca gtatcaagc 2220  
 acctgtgcct gtgcagggac catctaccg tggaaactgga atgaacaagat catcatctc 2280  
 gatattgata ggaatcaca caagtcgag atgtctggag agatctccc acagcttggc 2340  
 aaagactctg cccaacagga tatagcaaaag cctcaactc ccatcaatga gaaatctgca 2400  
 aagttctgta actgctcgcc tctgctcacc ggcattgccc acatgacccg tggctaccg 2460  
 cacttggttca atgaacaagg caaatctg ccaatctg cccctgacct gtcctccagc 2520  
 agctgttct cgcctcca cagaagaagt atagaagaag aaaccaagaa gtctcaaat 2580  
 gagtgtctaa atgatataa gaatctgtt gccctgtcta agcagccct ctatgctgc 2640  
 ttgtgaaacc gtccaatatg tglctatgcca taacacaaga ttgagttcc agactgtaga 2700  
 atatacccaa tgaaccccaa ggttgaaatc atacaagaag gaacccaagga aaaccaagtc 2760  
 tctgtacaga ggtcgttcca cctctcag cctctcag taaagagcag 2820  
 aatctccgct tccctgcc ggaattcagc tccctctgct actgctgaga cccgattccc 2880  
 gaagtggacc tggatggacc gtcttgaggg tctctgaggg tgggttgaggca gggcttggtc 2940  
 cccctccca cagcaaggga agcagctggt cctctctgct gaactcagat accagctcc 3000  
 cccagcgggg acgggtgctt cctgagctcc tccctctgct cctctctgct tcccagggc 3060  
 agctgctcag gctgcagagg tctgcagcgg aggaagaagg aggaagaagg aagaatctgt 3120  
 cctgcttgga actgcttgga acaatgttt accatgttt aggaatctgtga 3180  
 gctgggggct ccttccattg catcattta aagaagaagta aaagaagtg 3240  
 gaacaaaaca ctgcacaag tgaacacaag tgaagtlgtc aggttggctcc 3300  
 tggtaaccag gtctctca gatgtcaggg tctgctgccc cagtcaagtg 3360  
 tggctgctga agctgctggt gacagttt

<210> 86  
 <211> 1707  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1708229CB1

<400> 86  
 cctcgttca ccagccttg ccagcccttg cctcctgaaa ggggagcagc gttgtctctc 60  
 gccaagcgtc ccatctggcca ggaagtggag ccttcagggcc cttcccgctc cttcccgctc 120  
 ctgggagacc actgcagctc ccttgaaagc caagcccgcc ccttcagggcc ccttcagggcc 180  
 tgcctggcag ccgacacccc gcccgcctcc tctcccgctc cttcccgctc cttcccgctc 240  
 ggaactctgt gcaaaaatgt gcttcaaggc gttcagggcc tgcctcaaggaa gttccaaggcc 300  
 aagcccaatg gcaagaagcc cgttcgggag gaaaggaagg cctaacctgga gcttgagcac 360  
 accaaggcca ggatcacagg cttccagttc aaggagctgg tggctgctggc ccgctgagatc 420  
 gaccttaacg agtggctggc caggcaacac acgacgtttt tccaacacat caaccttgca 480  
 tatagcacca tctcggagtt ctcgacaagg gagaacgtgt agacgatggc cgtgtggcaac 540  
 accaagttac cccaagttc atgaggaagc agtgcacggc agtgcacggc agtgcacggc 600  
 gttgacctcg tcatgagctc cgttcagaaag cttggtgaagc atgaggaagc gttccccaca 660  
 aaatatcaga gagaatctcc cagctccctt gaagtccctt tgaaggaagat ctgcagacac 720  
 cgttccacg tcttgagaca catctactg gcccaactca aggaagacgct ggccttgag 780  
 ctgcacggac acttgaaacac gcttaacgct cacttcatcc cacttcatcc cacttcatcc 840  
 cctcgtgac





```
<210> 87
<211> 1752
<212> DNA
<213> Homo sapiens
```

87 <400>

[illegible]

```
<220>
<221> misc_feature
```



&lt;223&gt; Incyte ID No: 1806850CB1

&lt;400&gt; 88

ctgaagaagaa gattggagag tttaacaaga tttaacttcc gattttcttc ctcaacatgaa agaggaagaa  
 60 gtaggttttc agcccatgtt aatggaaatat tttaaccatg agagcatctc agagaaatctc  
 120 aagaaagtga ttgcacaaca ctgcctctcag aaggaatactg caagaaatctc  
 180 agcctatgga atcatgtctga agagcgcgaag aagtttctta aatatctcgt  
 240 tccatctgag tcaatctcgt aatatctcgt tcaatctcgt tcaatctcgt  
 300 cattgggcca gtaggtgacg gtaaaatatg gtaaaatatg gtaaaatatg  
 360 agcatgaaat ggtctcagct gacaaaaaac ggaatcgcct ggaatcgcct  
 420 agcatgaaat ggtctcagct gacaaaaaac ggaatcgcct ggaatcgcct  
 480 cattgggcca gtaggtgacg gtaaaatatg gtaaaatatg gtaaaatatg  
 540 gctgaacatg atgaatctga agagtctgag agagtctgag agagtctgag  
 600 gctgaacatg atgaatctga agagtctgag agagtctgag agagtctgag  
 660 atgtaaaaaa gtttaactca ttggtctaat ttggtctaat ttggtctaat  
 720 ttgaataacc ttatatttag ataacagtct gtaaaatatg gtaaaatatg  
 780 ttgaatcttc ttgaatcttc ttgaatcttc ttgaatcttc ttgaatcttc  
 840 gcatcttgaa gtttggcttc gtttggcttc gtttggcttc gtttggcttc  
 900 ggttcttgaa aatatcacaag ttgttgccct ttgttgccct ttgttgccct  
 960 aactctcatc aaagtgtgct ttgttgccct ttgttgccct ttgttgccct  
 1020 aaaaaataag acattacatc gtcagtccac aagcagttatg cctgttttga cgtatttaac  
 1080 aaacaagtga ttggagaaatg aatatgaatat ggaacacaccc ttgaatcttc  
 1140 agaatattca ctctcctta ttgtttgtag ttgtttgtag ttgtttgtag  
 1200 gatacttgag aatggagaca ttgtaatatg ttgtaatatg ttgtaatatg  
 1260 aaactttag ttgttgccct ttgtttgtag ttgtttgtag ttgtttgtag  
 1320 gtttcttgag agcagcatg ttgttgccct ttgtttgtag ttgtttgtag  
 1380 ttgtacagga cagctcttag aactatgta ttcaactccag aatctcttgc aatgtgttag  
 1440 aagaagcagaa ggaatgaatg gcttaaggag tttaactcttgc tttaactcttgc  
 1500 tctgtatcaag agacttgagc tttaactcttgc tttaactcttgc tttaactcttgc  
 1560 acaagaccag gttcagagtt ttgttgccct ttgtttgtag ttgtttgtag  
 1620 aactctctga atgtatgaat ctcttaactc ttgttgccct ttgtttgtag  
 1680 accgccaagt gtagccaaga ttgttgccct ttgtttgtag ttgtttgtag  
 1740 tgaacccctg ttgttgccct ttgtttgtag ttgtttgtag ttgtttgtag  
 1800 tgaacccctg ttgttgccct ttgtttgtag ttgtttgtag ttgtttgtag  
 1860 cttactcatc agcacatctc ttgttgccct ttgtttgtag ttgtttgtag  
 1920 cccaattctc acaactctag aatatctaat tttaactctc tttaactctc  
 1980 gcaaatatga ctcttggttc agaaaggttc tttaactctc tttaactctc  
 2040 tcaatttttc tttaaatga aatgctttaa agaatgttgc ttatgccaag tcaatttaag  
 2100 tatttttag aatatcttga gtttcttag tttaactctc tttaactctc  
 2160 taaacatgta ccaagcttgc cgaatcttgc tttaactctc tttaactctc  
 2220 aagaagcac attttgga aactaatctc tttaactctc tttaactctc  
 2280 tctcagaagg ccaattcaaa caaaccccaa tttaactctc tttaactctc  
 2340 ttgtctctga ttgttgccct tttaactctc tttaactctc tttaactctc  
 2400 ggttatcttc ttgttgccct tttaactctc tttaactctc tttaactctc  
 2461 taaccaaatc atttgagaa atgtlaatat aaaaatctac ttgaggaattc taaaaaataa

a

&lt;210&gt; 89

&lt;211&gt; 965

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1851534CB1

&lt;400&gt; 89

cttaagaaaa aaccctgtt gctgctgtla caaatlacaa caaatlacaa ggccttaaac  
 60 ggccttaaac ggccttaaac ggccttaaac ggccttaaac ggccttaaac  
 120 ggccttaaac ggccttaaac ggccttaaac ggccttaaac ggccttaaac  
 180 ggccttaaac ggccttaaac ggccttaaac ggccttaaac ggccttaaac  
 240 ggccttaaac ggccttaaac ggccttaaac ggccttaaac ggccttaaac  
 300 ggccttaaac ggccttaaac ggccttaaac ggccttaaac ggccttaaac  
 360 ggccttaaac ggccttaaac ggccttaaac ggccttaaac ggccttaaac  
 420 ggccttaaac ggccttaaac ggccttaaac ggccttaaac ggccttaaac



caaaccaaat accagtcata taatcaglat cctaattggt cagccaatgg ctttgtgca 480  
gtagaaact ttagcccaac tgaatcattat catcagaata ttccaacac aagaccacat 540  
gaaattcttg aaaaaccctc cctccacag ccaaccaccc ctccttcgt accacaact 600  
gtgattccaa agaatgctg ctcaaccctg attaactaa aataactaa aactatccag 660  
aatgacagg aattgttga gtcctccct tgtgagacc tttaataga agtaccagca 720  
agttagcaca cgaattcaaa gcatgaaagc agaaaagaaa agaggaataa aagcaacaag 780  
catgactcat agaatctga agagcgcag tcccaaatc agaaaccaa gggaggaacc 900  
gaacaaata tgaacaatg agaggttga cactgtatca gaaaaccaa acaaaggtag 960  
tgcc

<210> 90

<211> 2555

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Inyte ID No: 1868749CB1

<400> 90

agcacgtccc actctatgac cagtggagag atgtgatgaa agggatgaag gtggagtgc 60  
tcaacagtga tgcgtgtctc ccagccgggt tgtacttggat cgcctctgtc atccaagacag 120  
caggttatcg ggtgtcgtct cgttatgaag gctttgaaaa tgaacgccagc catgacattc 180  
ggtagaacct gggaacagtg gatgtccacc ccaattgctg gttgtgccatc aacagccaaga 240  
tcttagtacc atccaatgcaa agttccacga atctccacga atctccacga gttggagagca 300  
aacggtctgt gggctccagag acgcttcccg ttgatctcca catcagaatg catcagaatg 360  
tgaaatgccc ctttaggcag ggcattgcgc ttgaaagtgt ggaacaagtcc caggtgtcac 420  
gcactcgcac ggcctgtgtg gacacagtaa tccgtgtgctc acatgttgag cctccagagg 480  
atgtgtacag tgaacgacac tctgtgtgcc acatgttgag cccctctgat caccagtggt 540  
gttgttcacg acgttgtggc caggttgtga agatgtcacga atgtgtcacg gaaagcctga 600  
atccaccaca cttccacgttg atctaacctgtg atccacgttcc gaaagccttc aagaagtgac 660  
gagcactcta gctctctgtg ccagcagaaac ttgtggaaag gctgaaagggc actaaaggcc 720  
caaaagaaag ttgtggaaag aaagaaagaa aatcccgccc attaagagcc actaaaggcc 780  
gaacccctcc agagcctgtc aaagaaaggg cgcctgctcc gaaagagcac ttgtgtctcc 840  
agatctctgc gtagcctgtc cctggtcgtc atctgtctcc gaaagagcac ttgtgtctcc 900  
tgaacgtgtg aaacagacag aagcttccaa agtccctcaa gtcctgtctcc gtccagagcc 960  
tgaacgtgtg aaacagacag aagcttccaa agtccctcaa gtcctgtctcc gtccagagcc 1020  
agcctgagct gtaggtcgtg gacctgtgtg agcctcccggtc catctgtgtg gccacgggtga 1080  
aacgagtgtt gcatcgtctc ctaggcatac acttgcagc cttggtgacag gagtacgacc 1140  
agttgtgttg acgtcgtgtg ccagcagatc acccgtctcg cttgtgtgag ctcaaccggt 1200  
accagctcca gctctctgtg gctcctctgt gctcctctgt aatcccgccc actaaaggcc 1260  
caaaagaaag ttgtggaaag aaagaaagaa aatcccgccc actaaaggcc actaaaggcc 1320  
gaacccctcc agagcctgtc aaagaaaggg tgcctgctga tcatctgtct gtcctgtctcc 1380  
agatctctgc gtagcctgtc cctggtcgtc aagcttccaa gtcctgtctcc gtcctgtctcc 1440  
tgaacgtgtg aaacagacag aagcttccaa agtccctcaa gtcctgtctcc gtccagagcc 1500  
tcaagcagga aaacagacag aagcttccaa agtccctcaa gtcctgtctcc cttggtgagcc 1560  
agcctgagct tctcttaca ccaacacat gctccacat gctccacat gctccacat 1620  
tctctctgtt gtaaatctg cccgtgtctg tgaagcctg acgtgtgagag acctgtctgg 1680  
gtctctctgt acccctctgt tgcctctctg cctccctctg gaaaagtctc atgacagggc 1740  
cgctctaggg ccaggaactc gtcctgtgac gctcctctcc agccttctcc agcccaaggtt 1800  
gaaacgttagc tgggtccccc tgcctgtctc tgcctgtctc tgcctgtctc cgcctccccc 1860  
tgcctagggc acgtgtccac tgcctgtctc tgcctgtctc tgcctgtctc tgcctgtctc 1920  
cctgaacagag tcatgtcaat taagtcttag agcagctctc tgaagcctctc agagcctctc 1980  
gacagttagt ttgtgtgttg ggtcagctcc tgcctcaaaa atccaccaa atccaccaa 2040  
ctcagctcca tgtgtctctc cctgtctctc cctgtctctc cctgtctctc caggtatgtt 2100  
cctgtctctg cagctcaaga tcttgctgac ttctggcggc ttctggcggc ttctggcggc 2160  
cctgtagctg aaatgtgccc aaatgtgccc aaatgtgccc aaatgtgccc aaatgtgccc 2220  
ggggaggaaca ttggagggaa gatgtccctg ttgtgtccat ttgtgtccat ttgtgtccat 2280  
gaaagccccc taaaaaaat tcaatccaaa ttctcttctg ttctcttctg ttctcttctg 2340  
ctgtagctctc tagtagctg gctctgtatg gctctgtatg tctctgtatg tctctgtatg 2400  
aaagaaaccat gcttgaggg gcttgaggg gcttgaggg gcttgaggg gcttgaggg 2460  
ggaggaacaca tctagctgtcc atctgcaacc ttgcaacc ttgcaacc ttgcaacc 2520



2555

10

172

VNA

como sapientia

 $\leq 0$ 

misc-feature

DocId: 34262211

14

**£6/6L**





<210> 92  
<211> 4037  
<212> DNA  
<213> Homo sapiens

[illegible]



93 <400>

81/93



agcagaagct gagagtaaac tgaatacaagt atataaccac accatacagca aaccacaatcc 1080  
 caaccaataa tccactagta atgaggaaagc tggctgctgtg aatggagcttg atggagaccac 1140  
 gttccagcct cagaatctcc tcttagggag agcctgttag agctgctag ctacacagtc 1200  
 tccacagtg tatctctgg gcccaaccaa tatgcagtg agatlatgt caattgttg 1260  
 gcttattg aaaaaatag gaggttgaa aatgccacc cagtccaagaag aagaaagatc 1320  
 atctccagc ccaactacag aggaaccctc tgttagaagt cagtgtccc gccagggcat 1380  
 gacgtccgaa acacttccgaa acacttggag tccaaagtc gcagttgaaga cccgccagc 1440  
 ttctctct catatcac atttccaaca atttcagca caggctcga caggctcga 1500  
 ccggtctgag cagcagcaga gacggccgt tgtgtccat aatatgctg ccatlaggc 1560  
 agaatgttag atgctttaa atctctaac ttatatgtg ttgttgcga taacaataag 1680  
 ttctctct ttctctct ttgttgcga ttgttgcga ttgttgcga ttgttgcga 1740  
 cagcttca ttgttat tttaacat ttcttctg ttgttgcga ttgttgcga 1800  
 caacctggaa aaaaaaaat caaaacatg aaactctgt actctaac agagtagt 1860  
 gcttagcaaa agatgtgtg gagtgtgac tatccatgg ggttctgga tggaattgccc 1920  
 tgcagagccc ttatgcagc attttacc tttagtagt gccacaatg aaccccaag 1980  
 gatgctgta taatgagat ccatatcga gacagtaacg tccagctta catgattca 2031  
 ttaggttaa ataaattg ccaattaca ctaaaaaaa aaaaaaaaaa a

<210> 94  
 <211> 820  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Inyte ID No: 2456494CB1

<400> 94

aagggctcc gttagaacgc attcctgc aggcctccgt gcgcacctg gcgcacctg gcgcacctg 60  
 accgagccga gtcctgtcct tccagggcgt tgcgaatgtg ggaatgagtg ctcatggaac 120  
 tgcacgcgt ccttagtcgga caccggccc tgaatcaga gaaacagcag ctcatggaac 180  
 agctgtggtg gctgtgtgtg gagagggcca gctgtgtgc ccaagtacgt ccgctgagt 240  
 gccgggtgc cttcccgcaa acgtttaatg gcgaagatcc ccggtctccc gagtttatcg 300  
 tgcagacggc gtcctacatg ctgttgaag agaaccgat ctgcaaacga gccatgagc 360  
 tggcatctc aatcagcctc ctcaaccggg aagccgaagg gttgggtgtg cctacatcg 420  
 agatgtagat cccatctca gttgatatc gggcctctc cgtatgagatg aaacagttgt 480  
 ttggtctgga tgaagacgaa gacgaacgag acgaagaaag gtaggatg atctagggcc 540  
 tgcacctcg ggcctcgggg ggagggggcc tgcacggccc caaccctcc ccgagagcct 600  
 caccggcca gagcccatg ctctccctc ttcccttaac taaccggccc 660  
 cgtctgtc ctctctcat ttcccgtag tgcctgtc ttgtccagga atagcgtcc 720  
 agttacctg tgcctgtgag gggtctgga cctacacac tccgaaatgc tggaaagtgt 780  
 catctacct ggcctcccc gggtctccct cctgtcaat cctgaaatgc tggaaagtgt 820

<210> 95  
 <211> 2070  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Inyte ID No: 2668536CB1

<220>  
 <221> unsure  
 <222> 2058, 2067  
 <223> a, t, c, g, or other

<400> 95

gatggccga gtcggcaagg agaacgtcg ctgaagggt tgcctgaagc gaggggattc 60  
 taacatttc aggaaacct ttggaagaa caagcctac tcaataatg aaggagata 120  
 aagaaatc aagcctca gtaactcag caaacctgga ccaacaaga ccatgtgtg 180  
 actggataa gaagacttg gctacatac cctcaacaat tgaaggaat gatccagca 240  
 ccgaagccc gtaaccgga gagggcgtc ggttcattc tgaatgtggc acaagtttg 300



ggtacacata tgaataccctg gcaacatgga taattatct tcatcgcttc tatatgttcc 360  
atctctcaa gcaatctcaa agaatatgta caggagctg ttgcctcttc aatgctctgc 420  
aagttaaaa aatgtaaa aatgtaaa aatgtaaa aatgtaaa aatgtaaa 480  
atgatgtaca atttgccag ttgtgagatg acccaagga ggaagttaag 540  
gaaatctaac atttgagag ttgtgagatg aatgtaaa aatgtaaa aatgtaaa 600  
gaatctaac gcaagacatc aagttaaa aatgtaaa aatgtaaa aatgtaaa 660  
taaaatgac taagatgta aagaatgta aagaatgta aagaatgta aagaatgta 720  
tcaatgacat agcatgtatg tatctcgag gacgttctg caaatctgaa atcaagaat 780  
ggacaccaa acccatctgc caccatctgc tgaatctgc tgaatctgc tgaatctgc 840  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 900  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 960  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1020  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1080  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1140  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1200  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1260  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1320  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1380  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1440  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1500  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1560  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1620  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1680  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1740  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1800  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1860  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 1920  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 2040  
tgcctctgc caccatctgc caccatctgc caccatctgc caccatctgc caccatctgc 2070

&lt;210&gt; 96

&lt;211&gt; 2046

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Inctyle ID No : 2683225CB1

&lt;400&gt; 96

tgaacatgga tgggatatga tccctgggta acactctct tgcataagga tgaattctga 60  
ccaacttgaa tgcctatgat acaagaataa agatatctga agatatctga agatatctga 120  
gttgaaacca attctgtgga tagagaaact gttgaaactg ctttaacaag tatcttaaca 180  
ggcacaacca gtaatatgac cataagagc tacaagagc tacaagagc tacaagagc 240  
agccaagata cataagagat gacaatctctg aactcatatg gctcagaaa ctgatacaatc 300  
ctgttaaaaac ctcctgggaac catcaagatc tcaacagaga acrtcttatg aatcaaaaaa 360  
gggtgtctgc tctctagaaac aaaccaagat tgcagaaagt gatgaaagaa agaaacagag 420  
accaagtaat aagaagcaag gaaagcaagaa caagaagaa gaaatctgac ttggaataag 480  
agctataaaa acggcagcag aagtctggag agcttgaact tgaagaaact tgaagaaact 540  
aagaagcccc aagtctgga gaaagcaag caatctcag agaaacagcc tgcctgtctac 600  
aagaagcccc aagaagcccc agtctcag ctgaagcccc cgtgtgtctac 660  
ccaagagc tgcctgtgga atctctctca aatctcagag aaaaagcccc 720  
tcagccagca gtagaatctg cctgtgtctg agtatctggt tgaattctct acgatctggg 780  
tggatgtcctt gctgtgtgaa gttgcagttc ctatctcgca aatgaaagggc agtgcctccg 840  
acgttaagtt gaaatgtaga cctgtgtctg gaaagcaatc agaaaccaa agcaaaaaaa 900  
gttgaaagca gaaagagaa tgcctgggac cagagcccc tttctcttaa aagctaaaaa 1020  
ttttctctc attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1080  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1140  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1200  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1260  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1320  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1380  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1440  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1500  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1560  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1620  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1680  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1740  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1800  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1860  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 1920  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 2040  
aagctaaaaa attcceaatt tctgtggacc cagagcccc tttctcttaa aagctaaaaa 2070





aagaaaggtac atcctgtcca tccatcaat gcttctgatat ctagagaggtc tctctgtccag 1380  
tttctgtttt aaatgtcttg ctgatatctagt tcttctcagat ggaataaacct tccagttccct 1440  
ttagaggttg aaatagttcca tataaacctagc aaaggttagagc gcccctcagaa 1500  
ctttctcatc ctccaagagc agagtgtggga aggtttcccat gaccagcttg 1560  
tttctcttag catctcaatg tgaataatagt gtagacttgct gtccaagggagc ctctcatcaga 1620  
agatgtatatg catcttgaaag tctaaatgatat atctgatatcca agcttctgaaa 1680  
atctctgata aatgtctcatg tatctctta tcttctgttt tctctgtgaa gaaagacttt 1740  
caccaactgc ttagtgtatga tctgttgtat aaggtatgtag tctgatatcta ctatatgca 1800  
tctcagctctc cagaagaggg gctaggtcac agacagtagc tctctctgtt tctagcttagc 1860  
gttttaacct gctagtgttg gcatctggtac ctttgccttg ggtgtatatccg aagaaattgtt 1980  
gaggaatttag ctcata 2040  
2046

<210> 97

<211> 2660

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Inyte ID No: 2797839CBI

<400> 97

gttgcgtagtg ccgaccgaaa gctaggtccg gattgcacgt ggauggcgcg gtaguggcac 60  
ctctcggaca ttaaccgcga tctgttaca tgggcgcgaa gttggaacct gtaguggcgcg 120  
agcgggggag aggcggaaag gcccggaaag agaaaggttg ctagatgtgc ctagatgtgc 180  
tcttccctgc agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 240  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 300  
aaccatggcc tggaaagctca ccaaaagtgga tctctgtcag ccttaagac aaataagttc 360  
aagaaaggacc ccagtcccta tttaatgtct ctcgaagcca gaagcgcgca gcaacctgtgca 420  
gtgatgtgga agagagagag gaagactctg aagaaagatg tatgtgtgaa ccauggggacc 480  
tcttggtgctc cgaagagctat gctgtatcag ttagatgatat agctgtctcg aagcagaaag 540  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 600  
cccggaagc tgtctgtggc atccagtgga gtagaagaga gaccggagag aagcagaaag 660  
aagaaagagtg gaaaccttga tcaaggtccc ccaaggtlga agtagtgcag gtagcagtag 720  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 780  
ccaagctgcc agactctgca cgaattcaaa agcgttcaaa agcgttcaaa agcgttcaaa 840  
gtgatgtgga agtagtgcag gtagtgcag gtagtgcag gtagtgcag gtagtgcag 900  
ctctcggacc tggaggtlga tgaaggtlga tgaaggtlga tgaaggtlga tgaaggtlga 960  
1020  
tcttccctgc agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1080  
gttggtgttaa cctgtgtccc agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1140  
cttctgtgcc catgtgtgct acccggagtg acccggagtg acccggagtg acccggagtg 1200  
cctcagagag gttgtgagtg atgtgagtg acccggagtg acccggagtg acccggagtg 1260  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1320  
tgaatgtgga agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1380  
aacctgtgga agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1440  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1500  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1560  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1620  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1680  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1740  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1800  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1860  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1920  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 1980  
agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag agtagtgcag 2040  
2340



gggccaagg ggttgaggaag cagcagttgc cagagcagcc ttcttgagaa gctgcctcc 2400  
 agaaacagaa tgaataccccc aaggggctcc agcctccccc tgtgtctccc atcgtctcca 2460  
 gcccccac accagccaaag aggaagaaat ctccagttcc ggtgcacagc ggttgcaacc 2520  
 tatctatag tatctgaaac tagacgggtg gctcaatgcc atltgtcaacc ggttgcaacc 2580  
 ctgcctctg tgaagtagcc ttctctactg tgcataccca tgaatatcaa tacacatttt 2640  
 aaaaacctg aaaaataaaa

&lt;210&gt; 98

&lt;211&gt; 4610

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2959521CB1

&lt;400&gt; 98

ggggcccgga cgcattaggg ggtctcgccg ggtctcgccg gctgtctcac gctgagcagc 60  
 cctgcctccg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg 120  
 cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg 180  
 cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg 240  
 cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg 300  
 cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg cctgcggcgg 360  
 tgcgtagagt gacagagctca cctgcagaga gctgaagggc cctgcctcaacc acccttgagg 360  
 gggagtgggc cgggtgtlaag tcaacagccct tcaacagccct tcaacagccct tcaacagccct 420  
 ccttagttct ccttagttct ccttagttct ccttagttct ccttagttct ccttagttct 480  
 cagcagagctc cagcagagctc cagcagagctc cagcagagctc cagcagagctc cagcagagctc 540  
 cagcagagctc cagcagagctc cagcagagctc cagcagagctc cagcagagctc cagcagagctc 600  
 ctcgagagctc tccctctcgt gacccgtcac tccctctcgt tccctctcgt tccctctcgt 660  
 tccctctcgt tccctctcgt tccctctcgt tccctctcgt tccctctcgt tccctctcgt 720  
 cggagggcatt cagctcattc cagctcattc cagctcattc cagctcattc cagctcattc 780  
 tctgggggtct tctgggggtct tctgggggtct tctgggggtct tctgggggtct tctgggggtct 840  
 aaaaatgagc aaaaatgagc aaaaatgagc aaaaatgagc aaaaatgagc aaaaatgagc 900  
 tcccaagagg tcccaagagg tcccaagagg tcccaagagg tcccaagagg tcccaagagg 960  
 agcattcctc agcattcctc agcattcctc agcattcctc agcattcctc agcattcctc 1020  
 gctcgtctcc gctcgtctcc gctcgtctcc gctcgtctcc gctcgtctcc gctcgtctcc 1080  
 tccgcttacc tccgcttacc tccgcttacc tccgcttacc tccgcttacc tccgcttacc 1140  
 ggtgagagct ggtgagagct ggtgagagct ggtgagagct ggtgagagct ggtgagagct 1200  
 gcaacctcgt gcaacctcgt gcaacctcgt gcaacctcgt gcaacctcgt gcaacctcgt 1260  
 tctgtctcac tctgtctcac tctgtctcac tctgtctcac tctgtctcac tctgtctcac 1320  
 tggcctcaac tggcctcaac tggcctcaac tggcctcaac tggcctcaac tggcctcaac 1380  
 ggtctcgtcc ggtctcgtcc ggtctcgtcc ggtctcgtcc ggtctcgtcc ggtctcgtcc 1440  
 tgaattgagg tgaattgagg tgaattgagg tgaattgagg tgaattgagg tgaattgagg 1500  
 agaaccttgg ttatctacca agaaccttgg ttatctacca agaaccttgg ttatctacca 1560  
 agaaccttgg ttatctacca agaaccttgg ttatctacca agaaccttgg ttatctacca 1620  
 cctcctgact cctcctgact cctcctgact cctcctgact cctcctgact cctcctgact 1680  
 tgggttaggg tgggttaggg tgggttaggg tgggttaggg tgggttaggg tgggttaggg 1740  
 cccagcccgag gacttaacat cccagcccgag gacttaacat cccagcccgag gacttaacat 1800  
 acaagagctg tatcctcctga tcaaatgttc aaagtgttc aaagtgttc aaagtgttc 1860  
 taagctctggc cctcctgact tgaattgagc tgaattgagc tgaattgagc tgaattgagc 1920  
 ttaaccttga gaaagtgttc gaaagtgttc gaaagtgttc gaaagtgttc gaaagtgttc 1980  
 tcaaatgttc gaaagtgttc gaaagtgttc gaaagtgttc gaaagtgttc gaaagtgttc 2040  
 agagcagagg accctcagct agagcagagg accctcagct agagcagagg accctcagct 2100  
 cctcctcaagc cctcctcaagc cctcctcaagc cctcctcaagc cctcctcaagc cctcctcaagc 2160  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2220  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2280  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2340  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2400  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2460  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2520  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2580  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2640  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2700  
 gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt gaaacctgga tcaagagtggt 2760



aatggtctgc ctaatgata tgcatacagat gggtttaaa tgaaccgtct aggtatactgc 2820  
 tcccttgcaa aaaaagtgca atccttgcat gaattgaata tgaattcttc taactctctc 2880  
 cagaaaatgg atggagataa ctgtctctaa aaactgttagg ccagctctaa ccaacttgga 2940  
 gctcgtctgc tatcatatgg ataatgttagg cacttttggt cacttttggt agcctgtctc 3060  
 tccagaagcc tctcttagtg gttgccacag ttggagccca ttggcaaaatg 3120  
 atcaltgtc atgtctga caagtactgg ttcaactaca atgcttagc ttctctcta 3180  
 catagaaaaa ctgtccgtcc tcagtataca caagcagcat ccgttttgtt tctctctct 3240  
 gggaagatac tgaataaaca ggaatattct tgaataaaca gttgagcagga gttgagcagga 3300  
 gttgaattt tttaagttt tgttttatc ttgcctgttg gttcaatatc atttgagatc 3360  
 agctgaaga gggaataatt cagtgtatga gatcttagat taaatatcag gactgatttc 3420  
 ctgtgtggag gattatgttc cagtattacc aaagaaccaa tctcttgaa ttgtggaattc 3480  
 aacttttat atgtcatata ttattgttgt tttaaacgg ttctttgtct ttctgtttt 3540  
 attttctca agctgtcttc aggaagctagc agaaaataaac tcaagttga agactctgga 3600  
 agatttttgt taaacctaac tgcatttgat gttatataa ttataattta gcaattccaa 3660  
 tagattctat catctcttaa acataatatc ctgtcttg ggttagata ctaagttaga 3720  
 gttagttgat tcttagtta gtagagtagg tcaaaaactat aatcttaac aaattgaaaa 3780  
 atgaatatag gttgtttccc tttrtttgca cacttatat accrtaaaga atttctctc 3840  
 atagaacagt gctcaaaagg gaaatctct ttaaacctc ttgagcagga aggtcagttc 3900  
 tagtcggagc ttgagagggc ctagagagct cacatcgtct gacttgatc gccactgat 3960  
 gttgccaag cttgtcccca tgaagtcata atltgcaat tcttrttgat ttragtgtt 4020  
 gaatttgtt cacccttgcc attagcagc tgcacctcat ttccagatg caccagctcc 4140  
 tatataag ttagcaagg aaagttagg caaggttagg caaggttagg 4200  
 gttctgtct cctctcact caccacggc aacagcttg tgcctgtct ttgcccca 4260  
 gttatttgt gttcagttc aaattgagc tatctctaac ttgtctctaa ccttggttt 4320  
 taaaagaag gctctctgt ttggttaggc taaagctga gttatagtag tccrtctca 4380  
 aagtagctgc gcatctact ttaaacaaag ttgtctgat ttgccaagga 4440  
 agttagat ttatctgtc ttatctctc ttaagttct gcagttccat ccaagtattt 4500  
 tttaatat ctaggtgta tgaagaagaa ttgaataaag aaattaaact atgtggactg 4560  
 taaatgttt atttgtaaga ttctataat aaagctatat tctgtaaaac tctgtaaaac 4610

&lt;210&gt; 99

&lt;211&gt; 1889

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Inyte ID No: 3082014CB1

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; 1809, 1848, 1853, 1880

&lt;223&gt; a, t, c, g, or other

&lt;400&gt; 99

acgtagaaga tggagagaag gagggtgtat tgaagatatc tgaagatatc tgaagatatc 60  
 ctacacgggg ttctccaatc tctctctat agggggaat gcagttgag cctctgaa 120  
 gagacagata tcttagacca gtaacctgaa aagagagact gctttttta ctcaactgtg 180  
 tttagaccac actctccgct gatcagggcg agattagagt tgggttgcaaa tgglttgcaa 240  
 taccagcttg agatcccaag tcccttagta gaggtagaat ctgataatcg gaaaccaagc 300  
 aagatgtgaga tgaaggtctg ggaaccagag aacctctca cagacccgca gatccgaagc 360  
 ttctctgttg ttgcccagag ttgtgggaac ttgtcgaagc ccttagatg ttagcagctcc 420  
 atctgcagc atagcttgca catgagttga gttcgtctc cccgagatat cactctgttt 480  
 cagccatgg ataccttgca aaggaacggc taccgacctg ctaaggccat gtcgaacctg 540  
 gtaacccagg gaggccccgt gctgtgtcgg gatgagattg aggaattgtc agcctcagag 600  
 gccatgtcat ttgagtaggc cctagagaa gattggaaag actcaatga tatctgcag 660  
 gatttctac cctggaagtc acttgccagc atagtccagt ttatatcat ttggaaaaa 720  
 acgaacccgt atattcagca gaaaagtttg aaagcttgct aagcagagag caaaactgaaa 780  
 cagttctaca tcccaacta cactaagcca aagggccctga aggttgagag ttgcccacac 900  
 acaacgtctg ctcaagtgtat tgcctggggc tggattctca aagggccctga tgcagtgccg 960



<210> 100  
<211> 2032  
<212> DNA  
<213> Homo sapiens

[illegible]





```

<211> 1356
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<223> Incyte ID No: 4184320CB1
<400> 101
aatgaaagca acaaggagctg ctccggggac tgccttgc agtaccagga atcagtgctc 60
aggctcagaa atccttgata gaaaggagcat ttataaaag agaatgtgtc cacatcatatc 120
ccagcaacca agaccccac agtgttgtct ctccagatg agaaaaatga aagtcggctc 180
tcaccccag tatcctccgtg ctccagaaatg agaaaaatga aagtcggctc tcccgaaatg 240
acatccagtc tgaaaagtgg tccatcagca aacacacatca actcagccct accgatat 300
ttgggaacct tagttccaa gtaggttggc atcccaaca agcatagtat gtgcgatat 360
ctttgatcac aaaaacctgat ctccctac acctgatgac caaggaatgg cagtgtgagc 420
tcaccaagct tctcatctct gtccatgggg gccctgcagaa ctttgaactc cagccaatac 480
tcaagcaagt ctttgggaaa ggtctcatca aagcagctat gacaactgga gctgtgatat 540
tcaatggagg ggttaacaca ggtgttatlc gcatgtttgg cgatgccttg aaggatcatg 600
cctctaatgc tcatggcaag ataatgcccc tccgttatrg cccctggggg attgtggaaa 660
accaggaaga cctcattgga agaatgtttg tccggccatg tccagccatg tccaatccca 720
tgagcaagct cactgttctc aacagcatgc atcccaact catcttgtct gacaacggga 780
ccactggaaa atatggagca gagggtgaaa ttcgaagca actggaaaag catatttcac 840
tccagaagat aaacaaga tgcttgcgt tttctctct tgactcccg cttgtttatc 900
catrttgggg tagttgccag ttagaactcg tcaaggtgtt cctgtgtgtg 960
cactcatagt ggaaggagga cccaatgtga tctcgatrgt ttggagta ccttcgagca 1020
ccctcccggt gcaagtgtt gctgttgatg gtagttgacg ggcatacggg acctcggctc 1080
ttggcatcaa atactcagaa gaaagcgggt aggtatctt ccaagccca ttgaagaaac 1140
ctaaagcctg tctggaaaag aggttatgag ttgatctagt ttctcagtgc tcaaccaaga 1200
cctcaatca aaacaagcta tgacaatat gtctaaaaa tgtctgtcat gggagggctg 1260
tgttgaaaga cagaagaaaca tatcttaat gtccgtgaa gttggaaatt ctatgaaagc 1320
tacacggata ataaaaaggg tgaaggaaa agagga
1356
<210> 102
<211> 580
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<223> Incyte ID No: 4764233CB1
<400> 102
cacaaagcag gcaaccgactt cagtgtgcac gtctccttga caccctgcctc agtgtgcacg 60
ttcacttgag atcttcctc cgaaccctt gcccacgttg tgaccgcttg gtagcttga 120
gagtgtagag ggcaggttcc agcgtctg actcttctc tccatactgag acgcagctta 180
taggtccgca ggcagttcct ccaaggaact gaaatagtag aatatagagt ggcgaaggag 240
atcaacatat aggcctagc caagaaagaa tttàcagcct cctgagctga ttgggcttat 300
gcttgaaccc actgatgaag agcctaaga agaaacaaca cccaataaaa gtctggaatc 360
tacacccgat cagaagagag aagatgatca ggttgcactt gattgctgga gatggatg 420
ggaaagcagc tatgtcagagc taatgtcagag aagacacttt aatggatgag aagctggtac 480
tgatgtcaag gggagatlc taccaaaagc aagacacttt aagcactttt aagctggtac 540
agggaaatca caggtttaaa ggaagataag taccaaaagc aagacacttt aagcactttt 580

```



<210> 104  
<211> 2257  
<212> DNA  
<213> Homo sapiens

60	lgccccctgtg	gaagcccaacc	ataaatata	aaagcccaaga	1560
120	agagccctgtg	atgcccacac	atgcccacac	agagccctgtg	1580
180	agagccctgtg	atgcccacac	atgcccacac	agagccctgtg	1600
240	aaatcatag	tttctctga	tttctctga	aaatcatag	1620
300	tctcgtctgtg	tctcgtcagt	tctcgtcagt	tctcgtctgtg	1640
360	ggttttgaaag	atcgtcaatg	atcgtcaatg	ggttttgaaag	1660
420	ctgtctgtgtg	ttgaatacta	ttgaatacta	ctgtctgtgtg	1680
480	ttcagtgtaa	ttcagtgtaa	ttcagtgtaa	ttcagtgtaa	1700
540	agctcgttgtg	accacacacg	accacacacg	agctcgttgtg	1720
600	acagtcctgca	gctctagctg	gctctagctg	acagtcctgca	1740
660	ggttgcctgca	acagtcctgt	acagtcctgt	ggttgcctgca	1760
720	gcaacccctga	gggaagtctg	gggaagtctg	gcaacccctga	1780
780	gtcatgtaca	agacacccaa	agacacccaa	gtcatgtaca	1800
840	aaacatatcaa	catagataga	catagataga	aaacatatcaa	1820
900	aaaggtcataa	aggaatctct	aggaatctct	aaaggtcataa	1840
960	ggtcgtataac	ctctcgtctg	ctctcgtctg	ggtcgtataac	1860
1020	aggtcagctg	tcccgagctg	tcccgagctg	aggtcagctg	1880
1080	tccctgcaga	gcaatgctg	gcaatgctg	tccctgcaga	1900
1140	ctcctccaga	catcacaaaa	catcacaaaa	ctcctccaga	1920
1200	ggtgtcgtgtg	ggtgtcgtgtg	ggtgtcgtgtg	ggtgtcgtgtg	1940
1260	ggtgtcgtgtg	ggtgtcgtgtg	ggtgtcgtgtg	ggtgtcgtgtg	1960
1320	ggaaggaacg	tctctcagtc	tctctcagtc	ggaaggaacg	1980
1380	tctcccaagt	gacacacccc	gacacacccc	tctcccaagt	2000
1440	gctctccctgt	ggagtggtgtg	ggagtggtgtg	gctctccctgt	2020
1500	gtcttgaaact	ttctctctctg	ttctctctctg	gtcttgaaact	2040
1560	aaagcccaaga	ataaatata	ataaatata	aaagcccaaga	2060



<210> 105

<211> 2550

<212> DNA

<213> Homo sapiens

 $\langle 220 \rangle$ 

<221> misc\_feature

<223> Incyte ID No: 5627029CB1

<400> 105

[illegible]



cagaagacc cctacatcgt gctcagcggc agcggcgaag gcatgaaagc gaacacatc 2460  
accaagtgac agggccctcc cggcccgccc ctgcctcaac ctctcatlta aataaagctc 2520  
cctcctlat tttcaaaaa aaaaaaaaaa 2550

<210> 106  
<211> 2566  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Inyte ID No: 5678487CB1

<400> 106

cggtctcgagg tgaagactac aactccgac gtcgaagcg agggccagtg agggccagtg ggtggaag 60  
gcccccaaga gctctgtcgg gattcttag ctcgccctgt acagccgccc agggccagtg ggtggaag 120  
gcccccaaga gctctgtcgg gattcttag ctcgccctgt acagccgccc gaggaaagctg gacattctcc 180  
caggttaactc ccagctcttg gccctagagt cgtgcattgc gaagtctccc gaggaaagctg gacattctcc 240  
cccttgagga gattcttggg cagtattcaac gtagtctccc gtagaacattgc gaggaaagctg gacattctcc 300  
ttccaccaat gacatgtgccc ctgaagaagag gctgattctac tatctgggcaac gattctggcaac gattctggcaac 360  
atcttctctt tagtccaccagt gtgggcaaatg aggaagccctgg ccaggaattac cccgaattac 420  
aacagagccca tgcgtttaat cagctcaaaag atttctgtcgg ccacaaagca gatbaagaaa 480  
gtgaagtlac tccgtcaaga agaaagaaaaa tgtcccccctt gagggtcattg gaacattgagg 540  
aaaccattat gccctacatg caccgaacctg tccattacat taattgaaccag tctcatatata 600  
ttcatctatt agagccagaa gtttaagttct gcaagggagga actctcttga atgaaaaata 660  
aaatacaagt agtttgtctt gaataacggag ggctcccgca acagctcaaaa tctcaaaagac 720  
aagatagggaa actgagggaa caaacactcc tggattgcatt cggaaacatg cacaaattct 780  
ggattacaaac aggttgaaagt tctgggttgg gctgaacctc caaagaacca ttctccatg 840  
acaattcagga ttttgcaaaa gctgcattcg ctggttgaag gcttgaacctg gtagaagctaa 900  
aaacttacta tgaaggaaaaa tgtgaaatg aggaattcca attgaagtct ttaggaaagc 960  
acttaagctga atatacagga acttgttaga atcttaagga gcaacttaag cataaagaa 1020  
ttcttcttgc tgcataatc tgcataatc acccaatca atgttctaat gcaaacatc gaaagctcgg 1080  
atgaagctgt tcttcccaa acccaatca atgttctaat gcaaacatc gaaagctcgg gaaagctcgg 1140  
ttaaagaag agtagactg atgtctgcac tagttctcgt aagtagcagc ttgycagata 1200  
cgcaagcaaat agaaagcaagt gcttatgaac aggttgaaac aggttgaaac aggttgaaac 1260  
aaagccaatt tgaaaaaaac aaggtcttaa tccagtltga ccagtltga aagtagcttg 1320  
aagtagcagc ggaataagatc ttgcattctc gcaaggaata agggccattg 1380  
agaaagacat gatgaaaaag gaataataca aagaaagggga gtaacattggtg tcaaaagtct 1440  
agattcagc tattaataa ctggaagaaa ttcaaaagcca gctggtctct cgggaaattg 1560  
atgttcacaaa ggtgtgttgg gaattgtcgt atctagcttga agcgaacctg taataaccaa 1620  
atgtaggcagc aaagtagcac agaaagtlca agcgaataac taacagtgat ctgaaata 1680  
aagatcagga aatagagaaa ttgaagataa aaacttgatga aagcgaacca caacttggaa 1740  
aggaagcagcc ctaggcagag ctaggcagag aggaagtlcgt gtagcttaaca gaaactta 1800  
gcaaaccttg caactcaacc gacattcaaaa agatagcatc cagcagagct 1860  
ttagcaagga agcaaatgga accaaagccc aaggtccctc aaggtccctc aaggtccctc 1920  
agaagataca gcaaatgtga gcccaaacctg acaaaaactg aaatgaaacag tatrtgttgc 1980  
tgaaactcca gaatacatc ttgaacaagt taagggaaagt tcaactcagt caagaaaaaa 2100  
aaacttgaaa aatctctcaa aaaaaccagt ctgaataatg tgaagaaatg tgaagaaatg 2160  
gtgtcccaaga ttgtagagta catgagagca ttgaagcaaa gctaaagcag gctgataagc 2220  
acagcccaagc caggtccctc agctctgtc agtccctcag caagcagaa cagctctccc 2280  
tggaagagga gggccctgtcgg gaagaggttgg acccgtctcgg gaagagagta cccagctatc 2340  
cacaaatcga ttgttgacct ggtattgaac ggtattgaac ggtattgaac caaggttgc 2400  
ttaacattcat ctggtttaga cttatattgc caaaccatgc ccaggttgc 2460  
aacccctcag cctgcagttgg ggtcagttgg ggtcagttgg ggtcagttgg ggtcagttgg 2520  
tcgggtctgga gctggagctt gactctagct gtagcagagct ccttgt





<220>  
<221> misc\_Feature  
<223> Incyte ID No: 5682976CB1

<400> 107  
gcttctctta ttttttaaa tgttctataa tgcatacaag actatagaac tatctgtctt 60  
atgacacttc gaagaagtrc ggtccgtggt ggtccgtggt cacaacctgt agtgaagaagtt aggaacaagt ccctcgggtt 180  
ggaggaccgt tcccggtcgt ggaaccttta agagaattta tccaagaatt 240  
aaaccttgaa gcaaccttta atgtgaaatt aatacaactt gttggaattg 300  
cactggagaa tatatttat ctggtcagaa tgaaccaaaa ttagttaatt 360  
caggagaaag gtttgacaa caattcgttc agggcaccga gcaaacatat ttagtgcaaa 420  
gttcttaact tgaacaatg ataaacagat tglattcctgc tctggagatg gagtataatt 480  
ttataccaac gttgagcaag atgcagaaac caacagaaac tgcacaattt 540  
tgaaactaac tatgagatta tgaactgtac caatgaacct tacaatttcc tctcctgtgg 600  
tgaagatrga actgttaggt ggtttagtat acgatcaaaa actagctgca caaagaagaa 660  
ttgtaaagat gatattttaa ttaactgttcg acgttgcctgc acgttgcctg ctatttgcct 720  
accaatacca tatlaacctg ctgttgtgtg tcttgacagc ttagtacaag tatatgactg 780  
gccaatrgct ggaacaagat ctacagggaa tlatgcaagt tlatgcaagt cgaaggacta ctggaattgt 840  
tgaccttccc atcttaataa taaagtccctgc ttaactctca gatatacat atcttcttga 900  
cagtgaagat gttcaagaga tctcgttagt ttaactctca gatatacat atcttcttga 960  
ccgaaagat gatcaagcac gagaacttaa aactccttct gctggaagaa gctggaagaa 1020  
gttggaacca ccaacagtta agcgttltga acttcgttgt gatltgtcag ataactgacc 1080  
cagagcaagg caggagagtg aacgagaacg agatggagag cagagttcca atgttgtcat 1140  
gatgcagaga atgtctgata tgttatcaag atgttltgaa gaaagcaagt gaaagcaagt 1200  
aagcaatrga ggaacagcca gatactcgaa aaaaacttgaac ttaacagatg tctgaggttc 1260  
aaacttctct acgttcccat caagtccctga tcttgaaagt tctcaatcc cacttgaaagt 1320  
agatactcca gctgaacaat tctctcagcc tctcaatcc agtgaaaagt caatggaagt 1380  
tcaattcgaa tcaatcccat tctcaatcc tctcaatcc agtgaaaagt tctgaatcc 1440  
cagtgaaacca aggcagcttg ttgaggtcatc tggacaacca acacatcatc agtcttgatc 1500  
tcccttctct gttgttaaca aacagctcgt atccatgtca cttgacgagc aacaggtata 1560  
caataatrga aagcttgagcc ccaaacagag gacaggttga ccagtlttaa gtttgacata 1620  
cagcacagaa ggaacacata aagcacata ttaacagatg tctgaagag 1680  
tatagcatca agttctagag gaattggag ccaattgcaaa tctgaggttc aggaagagat 1740  
ttctggtcca cagagctcca tgcaaacacc agaaagagac agtgaacaac aagctcctga 1800  
agaaatcatc gaagtattga caaaatatca ggaagtata tcttgcaagaa acccagttga 1860  
gaaacctatc aatatacaac aatcagataa gttcacagcc aagccatttg atcccaactc 1920  
aggaagatca aatgaacctc atcttgatcg ctcttgttgg gttccagaaag aatctgcttc 1980  
atctgaaaaa gccaagaaac cagaaactc agaaactc agaaactc agaaactc gttgcataca 2040  
tgaataaac accaatcttg agctcagaa gccaactggtg atgtcctgag 2100  
tgaaagaaac tccaaccagg acctctgctc tcaagcacca gatgaacag atgtgacagt 2160  
agtccgtatc ccaggttgcaa gttatcggagc agaaccttgt gataagacgt ctgctgttgc 2220  
ccgtattcag gagtcttca gacggagaaa gaaatggaaa gaaatggaaa aattggatat 2280  
ttctgaactc aagagccgct tagtataaaa gtttataaa ggtccatcga actccagga 2340  
aatgaataaa gaagccaat tctgggttgc tctgggttgc taaacttltg atgtcctgag 2400  
ccaacttct acctgggatc ggcacacctg tagaccttg agtccctg aagctgataa 2460  
tcaatgttgc aacttgacct gtttgacca atttagcc catctgcat 2520  
agattatgac ataaagatc gttcacact gaaacgaaat agaatltta accgaaaaac 2580  
tgcgtatgaa gttataactc gaaacgaaat catgcttgaa gaaactaga acaacctaac 2640  
agtccaagcc tcttcaatg tgaagatgt ggttcaact aatcatatc gaactgaccg 2700  
atgtgaggtg gacagatcag aaggtctcgg tcaagagaaat gaaaatlgagg atgaggaata 2760  
gttaactctt acagagcttc acttaaatc tctgaattt gtaaaagca ttctatcat 2820  
tttttctctt acagagcttc agtgcactt taaggttatg catgaaattg ggagatttga 2940  
taaaacaaaa ctacagaaat gttttaaata gttttaaata gttttaaata gttttaaata 3000  
atgcaaaagt caatatttcc caatatttcc caatatttcc caatatttcc 3022

<210> 108  
<211> 2787  
<212> DNA  
<213> Homo sapiens  
<220>



	7	8	9	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
2787	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2760	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2700	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2640	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2580	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2520	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2460	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2400	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2340	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2280	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2220	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2160	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2100	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
2040	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1980	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1920	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1860	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1800	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1740	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1680	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1620	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1560	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1500	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1440	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	
1380	a	b	c	d	e	f	g																						



CORRECTED VERSION

(19) World Intellectual Property Organization  
International Bureau(10) International Publication Number  
WO 01/07471 A2(43) International Publication Date  
1 February 2001 (01.02.2001)

PCT

(51) International Patent Classification: C07K 14/00

(21) International Application Number: PCT/US00/19948

(22) International Filing Date: 21 July 2000 (21.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/145,075  
8 September 1999 (08.09.1999) US  
60/164,647  
10 November 1999 (10.11.1999) US

(63) Related by continuation (CON) or continuation-in-part

60/145,075 (CIP) to earlier applications:  
US  
Filed on 21 July 1999 (21.07.1999)  
60/153,129 (CIP)  
8 September 1999 (08.09.1999)  
60/164,647 (CIP)  
10 November 1999 (10.11.1999)

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and  
(75) Inventors/Applicants (for US only): HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US); LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US); TANG, Y., Tom [CN/US]; 4230 Kanwick Court, San Jose, CA 95118 (US); YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US); AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US); YANG, Junming

(54) Title: CELL CYCLE AND PROLIFERATION PROTEINS

(57) Abstract: The invention provides human cell cycle and proliferation proteins (CCYPR) and polynucleotides which identify and encode CCYPR. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of CCYPR.

WO 01/07471 A2



(48) Date of publication of this corrected version: 17 May 2001  
(15) Information about Correction: see PCT Gazette No. 20/2001 of 17 May 2001, Section II  
For two-letter codes and other abbreviations, refer to the "Guide to the PCT Gazette" appearing at the beginning of each regular issue of the PCT Gazette.

Published: Without international search report and to be republished upon receipt of that report.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(74) Agents: HAMLET-COX, Diana et al., Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

[CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). SHAH, Purvi [IN/US]; 1608 Queen Charlotte Drive #5, Sunnyvale, CA 94087 (US).



## CELL CYCLE AND PROLIFERATION PROTEINS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of cell cycle and proliferation proteins and to the use of these sequences in the diagnosis, treatment, and prevention of immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer.

### BACKGROUND OF THE INVENTION

Cell division is the fundamental process by which all living things grow and reproduce. In unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms, while in multicellular species many rounds of cell division are required to replace cells lost by wear or by programmed cell death, and for cell differentiation to produce a new tissue or organ. Details of the cell division cycle may vary, but the basic process consists of three principal events. The first event, interphase, involves preparations for cell division, replication of the DNA, and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell cytoplasm. The sequence and timing of cell cycle transitions are under the control of the cell cycle regulation system which controls the process by positive or negative regulatory circuits at various check points.

Mitosis marks the end of interphase and concludes with the onset of cytokinesis. There are four stages in mitosis, occurring in the following order: prophase, metaphase, anaphase and telophase. Prophase includes the formation of bi-polar mitotic spindles, composed of microtubules and associated proteins such as dynein, which originate from polar mitotic centers. During metaphase, the nuclear material condenses and develops kinetochore fibers which aid in its physical attachment to the mitotic spindles. The ensuing movement of the nuclear material to opposite poles along the mitotic spindles occurs during anaphase. Telophase includes the disappearance of the mitotic spindles and kinetochore fibers from the nuclear material. Mitosis depends on the interaction of numerous proteins. For example, mutation studies in the *Drosophila melanogaster zw10* gene show a disruption in chromosome segregation. ZW10 protein appears to function at the kinetochore as a tension-sensing checkpoint during the onset of anaphase. ZW10 appears to have a direct role in the recruitment of dynein to the kinetochore, and, dynein's involvement in the coordination of chromosome separation at the onset of anaphase and/or poleward movement (Start, D.A. et al. (1998) J. Cell Biol. 142:763-774).

Regulated progression of the cell cycle depends on the integration of growth control pathways with the basic cell cycle machinery. Cell cycle regulators have been identified by selecting for human and yeast cDNAs that block or activate cell cycle arrest signals in the yeast mating pheromone pathway

when they are overexpressed. Known regulators include human CPR (cell cycle progression restoration) genes, such as CPR8 and CPR2, and yeast CDC (cell division control) genes, including CDC91, that block the arrest signals. The CPR genes express a variety of proteins including tumor suppressor binding proteins, chaparones, transcription factors, translation factors, and RNA-binding proteins (Edwards, M.C. et al. (1997) *Genetics* 147:1063-1076).

5 The human CDC protein, CDC23, is homologous to the *S. cerevisiae* protein CDC23 which functions in the transition from metaphase to anaphase as well as in the exit from mitosis (Zhao, N. et al. (1998) *Genomics* 53:184-190). The *C. elegans* gene *cullin-1* (*cull1*) is a negative regulator of the cell cycle. *cull1* regulates the G1 to S phase transition and *C. elegans* *cull1* mutants exhibit hyperplasia of all tissues through acceleration of this transition by overriding mitotic arrest. *cull1* is a member of a conserved gene family that spans *S. cerevisiae*, nematodes and humans (Kipreos, E.T. et al. (1996) *Cell* 85:929-839).

10 Several cell cycle transitions, including the entry and exit of a cell from mitosis, are dependent upon the activation and inhibition of cyclin-dependent kinases (Cdk). The Cdk are composed of a kinase subunit, Cdk, and an activating subunit, cyclin, in a complex that is subject to many levels of regulation. There appears to be a single Cdk in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* whereas mammals have a variety of specialized Cdk. Cyclins act by binding to and activating cyclin-dependent protein kinases which then phosphorylate and activate selected proteins involved in the mitotic process. The Cdk-cyclin complex is both positively and negatively regulated by phosphorylation, and by targeted degradation involving molecules such as CDC4 and CDC53. In addition, Cdk are further regulated by binding to inhibitors and other proteins such as Sic1 that modify their specificity or accessibility to regulators (Patra, D. and W.G. Dunphy (1996) *Genes Dev.* 10:1503-1515; and Mathias, N. et al. (1996) *Mol. Cell Biol.* 16:6634-6643).

20 **Reproduction**

The male and female reproductive systems are complex and involve many aspects of growth and development. The anatomy and physiology of the male and female reproductive systems are reviewed in Guyton, A.C. ((1991) *Textbook of Medical Physiology*, W.B. Saunders Co., Philadelphia PA, pp.899-928).

25 The male reproductive system includes the process of spermatogenesis, in which the sperm are formed. Male reproductive functions are regulated by various hormones. The hormones exert their effects on accessory sexual organs, and are involved in cellular metabolism, growth, and other bodily functions.

Spermatogenesis begins at puberty as a result of stimulation by gonadotropic hormones released from the anterior pituitary. Immature sperm (spermatogonia) undergo several mitotic cell



divisions before undergoing meiosis and full maturation. The testes secrete several male sex hormones. Testosterone, the most abundant, is essential for growth and division of the immature sperm, and for the masculine characteristics of the male body. Three other male sex hormones, gonadotropin-releasing hormone (GnRH), luteinizing hormone (LH), and follicle-stimulating hormone (FSH), control sexual function.

The uterus, ovaries, fallopian tubes, vagina, and breasts comprise the female reproductive system. The ovaries and uterus are the source of ova and the location of fetal development, respectively. The fallopian tubes and vagina are accessory organs attached to the top and bottom of the uterus, respectively. Both the uterus and ovaries have additional roles in the development and loss of reproductive capability during a female's lifetime. The primary role of the breasts is lactation. Multiple endocrine signals from the ovaries, uterus, pituitary, hypothalamus, adrenal glands, and other tissues coordinate reproduction and lactation. These signals vary during the monthly menstruation cycle and during the female's lifetime. Similarly, the sensitivity of reproductive organs to these endocrine signals varies during the female's lifetime.

A combination of positive and negative feedback to the ovaries, pituitary and hypothalamus glands controls physiologic changes during the monthly ovulation and endometrial cycles. The anterior pituitary secretes two major gonadotropin hormones, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), regulated by negative feedback of steroids, most notably by ovarian estradiol. If fertilization does not occur, estrogen and progesterone levels decrease. This sudden reduction of the ovarian hormones leads to menstruation, the desquamation of the endometrium.

Hormones further govern all the steps of pregnancy, parturition, lactation, and menopause. During pregnancy large quantities of human chorionic gonadotropin (hCG), estrogens, progesterone, and human chorionic somatomammotropin (hCS) are formed by the placenta. hCG, a glycoprotein similar to luteinizing hormone, stimulates the corpus luteum to continue producing more progesterone and estrogens, rather than to involute as occurs if the ovum is not fertilized. hCS is similar to growth hormone and is crucial for fetal nutrition.

The female breast also matures during pregnancy. Large amounts of estrogen secreted by the placenta trigger growth and branching of the breast milk ductal system while lactation is initiated by the secretion of prolactin by the pituitary gland.

Parturition involves several hormonal changes that increase uterine contractility toward the end of pregnancy, as follows. The levels of estrogens increase more than those of progesterone. Oxytocin is secreted by the neurohypophysis. Concomitantly, uterine sensitivity to oxytocin increases. The fetus itself secretes oxytocin, cortisol (from adrenal glands), and prostaglandins. Menopause occurs when most of the ovarian follicles have degenerated. The ovary then

produces less estradiol, reducing the negative feedback on the pituitary and hypothalamus glands. Mean levels of circulating FSH and LH increase, even as ovulatory cycles continue. Therefore, the ovary is less responsive to gonadotropins, and there is an increase in the time between menstrual cycles. Consequently, menstrual bleeding ceases, and reproductive capability ends.

**5 Differentiation and Proliferation**

Tissue growth involves complex and ordered patterns of cell proliferation, cell differentiation, and apoptosis. Cell proliferation must be regulated to maintain both the number of cells and their spatial organization. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, tumor-suppressor proteins, and mitosis-promoting factors.

Embryogenesis is a process in which distinct patterns of protein expression control proper development. This process involves a host of proteins each with distinct and highly coordinated expression patterns. For example, in the mouse, temporally regulated expression of two related *Misg1* and *Mrg1* contribute to normal embryonic development. *Misg1* is expressed in the posterior domains of the developing mesoderm, while *Mrg1* is expressed in the anterior visceral endoderm. Properly coordinated expression of each protein throughout embryogenesis is critical for proper tissue and organ formation (Dunwoodie, S.L. et al. (1998) *Mech. Dev.* 72:27-40).

Growth factors were originally described as serum factors required to promote cell proliferation. Most growth factors are large, secreted polypeptides that act on cells in their local environment. Growth factors bind to and activate specific cell surface receptors and initiate intracellular signal transduction cascades. Many growth factor receptors are classified as receptor tyrosine kinases which undergo autophosphorylation upon ligand binding. Autophosphorylation enables the receptor to interact with signal transduction proteins characterized by the presence of SH2 or SH3 domains (Src homology regions 2 or 3). These proteins then modulate the activity state of small G-proteins, such as Ras, Rab, and Rho, along with GTPase activating proteins (GAPs), guanine nucleotide releasing proteins (GNRPs), and other guanine nucleotide exchange factors. Small G proteins act as molecular switches that activate other downstream events, such as mitogen-activated protein kinase (MAP kinase) cascades. MAP kinases ultimately activate transcription of mitosis-promoting genes.

In addition to growth factors, small signaling peptides and hormones also influence cell proliferation. These molecules bind primarily to another class of receptor, the trimeric G-protein

coupled receptor (GPCR), found predominantly on the surface of immune, neuronal and neuroendocrine cells. Upon ligand binding, the GPCR activates a trimeric G protein which in turn triggers increased levels of intracellular second messengers such as phospholipase C,  $Ca^{2+}$ , and cyclic AMP. Most GPCR-mediated signaling pathways indirectly promote cell proliferation by causing the secretion or breakdown of other signaling molecules that have direct mitogenic effects. These signaling cascades often involve activation of kinases and phosphatases. Some growth factors, such as some members of the transforming growth factor beta (TGF- $\beta$ ) family, act on some cells to stimulate cell proliferation and on other cells to inhibit it. Growth factors may also stimulate a cell at one concentration and inhibit the same cell at another concentration. Most growth factors also have a multitude of other actions besides the regulation of cell growth and division: they can control the proliferation, survival, differentiation, migration, or function of cells depending on the circumstance. For example, the tumor necrosis factor/nerve growth factor (TNF/NGF) family can activate or inhibit cell death, as well as regulate proliferation and differentiation. The cell response depends on the type of cell, its stage of differentiation and transformation status, which surface receptors are stimulated, and the types of stimuli acting on the cell (Smith, A. et al. (1994) Cell 76:959-962; and Nocentini, G. et al. (1997) Proc. Natl. Acad. Sci. USA 94:6216-6221).

Neighboring cells in a tissue compete for growth factors, and when provided with "unlimited" quantities in a perfused system will grow to even higher cell densities before reaching density-dependent inhibition of cell division. Cells often demonstrate an anchorage dependence of cell division as well. This anchorage dependence may be associated with the formation of focal contacts linking the cytoskeleton with the extracellular matrix (ECM). The expression of ECM components can be stimulated by growth factors. For example, TGF- $\beta$  stimulates fibroblasts to produce a variety of ECM proteins, including fibronectin, collagen, and tenascin (Pearson, C.A. et al. (1988) EMBO J. 7:2977-2981). In fact, for some cell types, specific ECM molecules, such as laminin or fibronectin, may act as growth factors. Tenascin-C and -R, expressed in developing and lesioned neural tissue, provide stimulator/anti-adhesive or inhibitory properties, respectively, for axonal growth (Faissner, A. (1997) Cell Tissue Res. 290:331-341).

Cancers and immune disorders are characterized by uncoordinated cell proliferation. Cancers are associated with the activation of oncogenes which are derived from normal cellular genes. These oncogenes encode oncoproteins which convert normal cells into malignant cells. Some oncoproteins are mutant isoforms of the normal protein, and other oncoproteins are abnormally expressed with respect to location or amount of expression. The latter category of oncoprotein causes cancer by altering transcriptional control of cell proliferation. Five classes of oncoproteins are known to affect cell cycle controls. These classes include growth factors, growth factor receptors, intracellular signal

transducers, nuclear transcription factors, and cell-cycle control proteins. Viral oncogenes are integrated into the human genome after infection of human cells by certain viruses. Examples of viral oncogenes include v-src, v-abl, and v-fps. Certain cell proliferation disorders can be identified by changes in the protein complexes that normally control progression through the cell cycle. A primary treatment strategy involves reestablishing control over cell cycle progression by manipulation of the proteins involved in cell cycle regulation (Nigg, E.A. (1995) *BioEssays* 17:471-480).

Many oncogenes have been identified and characterized. These include sis, erbA, erbB, her-2, mutated G<sub>s</sub>, src, abl, ras, crk, jun, fos, myc, and mutated tumor-suppressor genes such as RB, p53, mdm2, Cip1, p16, and cyclin D. Transformation of normal genes to oncogenes may also occur by chromosomal translocation. The Philadelphia chromosome, characteristic of chronic myeloid leukemia and a subset of acute lymphoblastic leukemias, results from a reciprocal translocation between chromosomes 9 and 22 that moves a truncated portion of the proto-oncogene c-abl to the breakpoint cluster region (bcr) on chromosome 22.

Mutations which hyperactivate oncogenes result in cell proliferation. Stimulation of a cell by growth factors activates two sets of gene products, the early-response genes and the delayed-response genes. Early-response gene products include *myc*, *fos*, and *jun*, all of which encode gene regulatory proteins. These regulatory proteins lead to the transcriptional activation of a second set of genes, the delayed-response genes, which include the cell-cycle regulators Cdk and cyclins. For example, the human T-cell leukemia virus type I (HTLV-I) Tax transactivator protein acts as an early response gene by enhancing the activity of a cellular transcription factor. The oncogenic properties of the Tax protein include transformation of primary T-lymphocytes and fibroblasts through cooperation with the a GTP-binding protein, Ras. Recently investigators have shown that Tax interacts with several PDZ-containing proteins. The PDZ domain, originally described in the *Drosophila* tumor suppressor protein Discs-Large, is common to membrane proteins thought to be involved in clustering receptors in growth factor signal transduction pathways (Rousset, R. et al. (1998) *Oncogene* 16:643-654).

Tumor-suppressor genes are involved in regulating cell proliferation. Mutations which cause reduced or loss of function in tumor-suppressor genes result in uncontrolled cell proliferation. For example, the retinoblastoma gene product (RB), in a non-phosphorylated state, binds several early-response genes and suppresses their transcription, thus blocking cell division. Phosphorylation of RB causes it to dissociate from the genes, releasing the suppression, and allowing cell division to proceed. Other gene products involved in cell proliferation, differentiation, and apoptosis are yet to be discovered. One method currently being utilized to help identify such new molecules involves comparisons between quiescent and proliferative tissues. For example, a subtractive hybridization screen of human placental cytotrophoblast cells identified 20 genes whose expression levels rose due to

EGF induction of cell proliferation. (Morris, D.W. et al. (1996) *Placenta* 17:431-441). Another method involves identification of molecules produced in cells treated with anti-tumorigenic agents, such as dithiolethiones. Presumably, the protective action of these anti-tumorigenic agents is associated with the induction of tumor suppressor gene products (Primiano, T. et al. (1996) *Carcinogenesis* 17:2297-2303).

5 In another example, the candidate tumor-suppressor gene ING1, that codes a nuclear protein, p33ING1, is involved in the negative regulation of cell proliferation. The action of p33ING1 is dependent upon the activity of another tumor-suppressor gene, p53. p53 is a cellular stress-responsive gene requiring the activity of p33ING1 to effectively induce growth inhibition of cells. p33ING1 and p53 have been shown to physically associate through immunoprecipitation studies (Garikavisev, I. et al. (1998) *Nature* 391:295-298).

**Apoptosis**

Apoptosis is the genetically controlled process by which unneeded or defective cells undergo programmed cell death. Selective elimination of cells is as important for morphogenesis and tissue remodeling as is cell proliferation and differentiation. Lack of apoptosis may result in hyperplasia and other disorders associated with increased cell proliferation. Apoptosis is also a critical component of the immune response. Immune cells such as cytotoxic T-cells and natural killer cells prevent the spread of disease by inducing apoptosis in tumor cells and virus-infected cells. In addition, immune cells that fail to distinguish self molecules from foreign molecules must be eliminated by apoptosis to avoid an autoimmune response.

20 Apoptotic cells undergo distinct morphological changes. Hallmarks of apoptosis include cell shrinkage, nuclear and cytoplasmic condensation, and alterations in plasma membrane topology. Biochemically, apoptotic cells are characterized by increased intracellular calcium concentration, fragmentation of chromosomal DNA, and expression of novel cell surface components.

25 The molecular mechanisms of apoptosis are highly conserved, and many of the key protein regulators and effectors of apoptosis have been identified. Apoptosis generally proceeds in response to a signal which is transduced intracellularly and results in altered patterns of gene expression and protein activity. Signaling molecules such as hormones and cytokines are known both to stimulate and to inhibit apoptosis through interactions with cell surface receptors. Transcription factors also play an important role in the onset of apoptosis. A number of downstream effector molecules, particularly proteases such as the cysteine proteases called caspases, have been implicated in the degradation of cellular components and the proteolytic activation of other apoptotic effectors.

## Aging and Senescence

Studies of the aging process or senescence have shown a number of characteristic cellular and

molecular changes (Fauci, A.S. et al. (1998) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, p.37). These characteristics include increases in chromosome structural abnormalities, DNA cross-linking, incidence of single-stranded breaks in DNA, losses in DNA methylation, and degradation of telomere regions. In addition to these DNA changes, post-translational alterations of proteins increase including deamidation, oxidation, cross-linking, and nonenzymatic glycosylation. Still further molecular changes occur in the mitochondria of aging cells through deterioration of structure. These changes eventually contribute to decreased function in every organ of the body.

The discovery of new cell cycle and proliferation proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer.

## SUMMARY OF THE INVENTION

The invention features purified polypeptides, cell cycle and proliferation proteins, referred to collectively as "CCYPR" and individually as "CCYPR-1," "CCYPR-2," "CCYPR-3," "CCYPR-4," "CCYPR-5," "CCYPR-6," "CCYPR-7," "CCYPR-8," "CCYPR-9," "CCYPR-10," "CCYPR-11," "CCYPR-12," "CCYPR-13," "CCYPR-14," "CCYPR-15," "CCYPR-16," "CCYPR-17," "CCYPR-18," "CCYPR-19," "CCYPR-20," "CCYPR-21," "CCYPR-22," "CCYPR-23," "CCYPR-24," "CCYPR-25," "CCYPR-26," "CCYPR-27," "CCYPR-28," "CCYPR-29," "CCYPR-30," "CCYPR-31," "CCYPR-32," "CCYPR-33," "CCYPR-34," "CCYPR-35," "CCYPR-36," "CCYPR-37," "CCYPR-38," "CCYPR-39," "CCYPR-40," "CCYPR-41," "CCYPR-42," "CCYPR-43," "CCYPR-44," "CCYPR-45," "CCYPR-46," "CCYPR-47," "CCYPR-48," "CCYPR-49," "CCYPR-50," "CCYPR-51," "CCYPR-52," "CCYPR-53," "CCYPR-54." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-54.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54.

54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:55-108.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of

SEQ ID NO:55-108, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, c) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

5 Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides complementary to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

20 The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

30 The invention further provides a pharmaceutical composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid



sequence selected from the group consisting of SEQ ID NO:1-54, and a pharmaceutically acceptable excipient. In one embodiment, the pharmaceutical composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CCYPR, comprising administering to a patient in need of such comprising administering to a patient in need of such treatment the pharmaceutical composition.

5 The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CCYPR, comprising administering to a patient in need of such treatment the pharmaceutical composition.

20 Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CCYPR, comprising administering to a patient in need of such comprising administering to a patient in need of such treatment the pharmaceutical composition.

30 The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group

35

consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-54. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

20 The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:55-108, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

25 The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group

consisting of SEQ ID NO:55-108, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55-108, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of the above polynucleotide sequence; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

10

## BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOS), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding CCYPR.

15

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of CCYPR.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

20

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding CCYPR were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

25

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

30

It must be noted that as used herein and in the appended claims, the singular forms "a," "an,"

and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

35

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

- 10 "CCYP<sup>R</sup>" refers to the amino acid sequences of substantially purified CCYP<sup>R</sup> obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.
- The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CCYP<sup>R</sup>. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CCYP<sup>R</sup> either by directly interacting with CCYP<sup>R</sup> or by acting on components of the biological pathway in which CCYP<sup>R</sup> participates.
- 15 An "allelic variant" is an alternative form of the gene encoding CCYP<sup>R</sup>. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.
- 20 Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.
- "Altered" nucleic acid sequences encoding CCYP<sup>R</sup> include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CCYP<sup>R</sup> or a polypeptide with at least one functional characteristic of CCYP<sup>R</sup>. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CCYP<sup>R</sup>, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CCYP<sup>R</sup>.
- 25 The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CCYP<sup>R</sup>. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CCYP<sup>R</sup> is retained. For example, negatively charged amino
- 30

- acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophobicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophobicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.
- 5 The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.
- 10 "Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.
- The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CYP<sub>17</sub>. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CYP<sub>17</sub> either by directly interacting with CYP<sub>17</sub> or by acting on components of the biological pathway in which CYP<sub>17</sub> participates.
- 20 The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CYP<sub>17</sub> polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.
- The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.
- The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as
- 35

phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CYP<sub>R</sub>, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CYP<sub>R</sub> or fragments of CYP<sub>R</sub> may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino

acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation,

(b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide. A "fragment" is a unique portion of CCYPR or the polynucleotide encoding CCYPR which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment

used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:55-108 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:55-108, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:55-108 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:55-108 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:55-108 and the region of SEQ ID NO:55-108 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-54 is encoded by a fragment of SEQ ID NO:55-108. A fragment of SEQ ID NO:1-54 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-54. For example, a fragment of SEQ ID NO:1-54 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-54. The precise length of a fragment of SEQ ID NO:1-54 and the region of SEQ ID NO:1-54 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular



biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at

<http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gort/b12.html>. The

"BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62  
Reward for match: 1  
Penalty for mismatch: -2  
Open Gap: 5 and Extension Gap: 2 penalties  
Gap x drop-off: 50  
Expect: 10  
Word Size: 11  
Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

5 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

10 Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity"

15 between aligned polypeptide sequence pairs. Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*  
*Open Gap: 11 and Extension Gap: 1 penalties*  
*Gap x drop-off: 50*  
*Expect: 10*  
*Word Size: 3*  
*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain

DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely

5 resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific

hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e.,

10 binding between pairs of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and annealing of nucleic acid sequences among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive

annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

20 Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic

strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency

conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_{61}$  or  $R_{61}$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CCYPR which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CCYPR which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of CCYPR. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CCYPR.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which

comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an CCYPR may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CCYPR.

"Probe" refers to nucleic acid sequences encoding CCYPR, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al., 1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection

programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell. Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid.

amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding CCYPR, or fragments thereof, or CCYPR itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type

of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule.

Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at



least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human cell cycle and proliferation proteins (CCYPR), the polynucleotides encoding CCYPR, and the use of these compositions for the diagnosis, treatment, or prevention of immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding CCYPR. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOS) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each CCYPR were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each CCYPR and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis along with relevant citations, all of which are expressly incorporated by reference herein in their entirety; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding CCYPR. The first column of Table 3 lists the nucleotide

SEQ ID NOS. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:55-108 and to distinguish between SEQ ID NO:55-108 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue

categories which express CCYPR as a fraction of total tissues expressing CCYPR. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing CCYPR as a fraction of total tissues expressing CCYPR. Column 5 lists the vectors used to subclone each cDNA library. Of particular note is the expression of SEQ ID NO:66 in inflammatory tissues. It should be noted that SEQ ID NO:76 was found to be expressed predominantly in nervous tissue.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding CCYPR were isolated. Column 1 references the nucleotide SEQ ID NOS, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2. SEQ ID NO:61 maps to chromosome 5 within the interval from 141.40 to 142.60 centiMorgans. This interval also contains gene(s) and/or EST(s) associated with corneal dystrophy and deafness.

SEQ ID NO:73 maps to chromosome 2 within the interval from 73.80 to 83.50 centiMorgans. This interval also contains gene(s) and/or EST(s) associated with hereditary nonpolyposis colorectal carcinoma and Muir-Torre syndrome. SEQ ID NO:74 maps to chromosome 19 within the interval from 41.70 to 58.70 centiMorgans. SEQ ID NO:75 maps to chromosome 17 within the interval from 62.90 to 64.20 centiMorgans. This interval also contains gene(s) and/or EST(s) located within the human breast cancer (BRCA1) gene region. SEQ ID NO:76 maps to chromosome 1 within the interval from 143.30 to 153.90 centiMorgans, to chromosome 3 within the interval from 156.20 to 160.00 centiMorgans, and to chromosome X within the interval from 112.80 to 139.40 centiMorgans. The interval on chromosome X from 112.80 to 139.40 centiMorgans also contains gene(s) and/or EST(s) associated with X-linked agammaglobulinemia.

SEQ ID NO:77 maps to chromosome 23 within the interval from 173.60 to 179.80

centiMorgans, and to chromosome 11 within the interval from 136.90 centiMorgans to q-terminus. SEQ ID NO:78 maps to chromosome 3 within the interval from 200.00 to 213.70 centiMorgans.

SEQ ID NO:81 maps to chromosome 7 within the interval from 167.60 centiMorgans to q-terminus.

SEQ ID NO:90 maps to chromosome 2 within the interval from 236.10 to 240.20 centiMorgans, to

chromosome 3 within the interval from 16.50 to 43.00 centiMorgans, and to chromosome 6 within

the interval from 124.20 to 126.50 centiMorgans. SEQ ID NO:91 maps to chromosome 2 within the

interval from 22.40 to 40.70 centiMorgans. SEQ ID NO:98 maps to chromosome 8 within the

interval from 40.30 to 60.00 centiMorgans. SEQ ID NO:100 maps to chromosome 14 within the

interval from 95.50 to 103.70 centiMorgans, and to chromosome 6 within the interval from 158.50

centiMorgans to q-terminus. SEQ ID NO:104 maps to chromosome 18 within the interval from 32.40

to 42.70 centiMorgans. SEQ ID NO:105 maps to chromosome 19 within the interval from 69.90 to

81.20 centiMorgans.

The invention also encompasses CCYPR variants. A preferred CCYPR variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CCYPR amino acid sequence, and which contains at least one functional or structural characteristic of CCYPR.

The invention also encompasses polynucleotides which encode CCYPR. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:55-108, which encodes CCYPR. The polynucleotide sequences of SEQ ID NO:55-108, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding CCYPR. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CCYPR. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:55-108 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:55-108. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CCYPR.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CCYPR, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CCYPR, and all such variations are to be considered as being specifically disclosed. Although nucleotide sequences which encode CCYPR and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CCYPR under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CCYPR or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CCYPR and its derivatives without altering the encoded amino acid sequences

include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CCYPR and CCYPR derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CCYPR or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:55-108 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENCE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV),

PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) *Short Protocols in Molecular Biology*, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) *Molecular Biology and Biotechnology*, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CCYPR may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a

known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.)

Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

15 When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

20 Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

30 In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CCYPR may be cloned in recombinant DNA molecules that direct expression of CCYPR, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CCYPR.

The nucleotide sequences of the present invention can be engineered using methods generally

known in the art in order to alter CCYP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Crameri, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or improve the biological properties of CCYP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CCYP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, CCYP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of CCYP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.)

The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active CCYPR, the nucleotide sequences encoding CCYPR or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CCYPR. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CCYPR. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CCYPR and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CCYPR and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CCYPR. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu,

N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J. J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al., (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R. M. et al. (1985) Nature 317(6040):813-815; McGregor, D. P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I. M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CCYPR. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CCYPR can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CCYPR into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Hecke, G. and S. M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CCYPR are needed, e.g., for the production of antibodies, vectors which direct high level expression of CCYPR may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CCYPR. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, *supra*; and Score, *supra*.)

Plant systems may also be used for expression of CCYPR. Transcription of sequences encoding CCYPR may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, *supra*; Broglie, *supra*; and Winter, *supra*.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated



transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CYP<sub>R</sub> may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CYP<sub>R</sub> in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) For long term production of recombinant proteins in mammalian systems, stable expression of CYP<sub>R</sub> in cell lines is preferred. For example, sequences encoding CYP<sub>R</sub> can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the induced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apv* cells, respectively. (See, e.g., Wigler, M. et al. (1977) *Cell* 11:223-232; Lowy, I. et al. (1980) *Cell* 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to

chlorosulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci. USA* 77:3567-3570; Colbere-Garapin, F. et al. (1981) *J. Mol. Biol.* 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), B

glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

5 Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CCYPR is inserted within a marker gene sequence, transformed cells containing marker gene can be placed in tandem with a sequence encoding CCYPR under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CCYPR and that express CCYPR may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR

15 amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences. Immunological methods for detecting and measuring the expression of CCYPR using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence

20 activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CCYPR is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St. Paul MN, Sect. IV; Colligan, J.E. et al. (1997) *Current Protocols in Immunology*, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) *Immunochemical Protocols*, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CCYPR include

30 oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CCYPR, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega

(Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for case of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CCYPR may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CCYPR may be designed to contain signal sequences which direct secretion of CCYPR through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture

Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CCYPR may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CCYPR protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CCYPR activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose

binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoprecipitation of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CCYPR encoding sequence and the heterologous protein sequence, so that CCYPR may be cleaved away from the heterologous moiety following purification. Methods for fusion protein

expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially

available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CCYPs may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

CCYP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CCYP. At least one and up to a plurality of test compounds may be screened for specific binding to CCYP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of CCYP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, Colligan, J.E. et al. (1991) *Current Protocols in Immunology* 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which CCYP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express CCYP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing CCYP or cell membrane fractions which contain CCYP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CCYP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CCYP, either in solution or affixed to a solid support, and detecting the binding of CCYP to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

CCYP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CCYP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CCYP activity, wherein CCYP is combined with at least one test compound, and the activity of CCYP in the presence of a test compound is compared with the activity of CCYP in the absence of the test compound. A change in the activity of CCYP in the presence of the test compound is

indicative of a compound that modulates the activity of CCYP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising CCYP under conditions suitable for CCYP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CCYP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CCYP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecci, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CCYP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding CCYP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CCYP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CCYP, e.g., by secreting CCYP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CCYPR and cell cycle and proliferation proteins. In addition, the expression of CCYPR is closely associated with inflammation, trauma, cell proliferation and cancer. Therefore, CCYPR appears to play a role in immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer. In the treatment of disorders associated with increased CCYPR expression or activity, it is desirable to decrease the expression or activity of CCYPR. In the treatment of disorders associated with decreased CCYPR expression or activity, it is desirable to increase the expression or activity of CCYPR.

10 Therefore, in one embodiment, CCYPR or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR. Examples of such disorders include, but are not limited to, an immune disorder

such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxicins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia,

20 irritable bowel syndrome, mixed connective tissue disorder (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner

25 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucocutaneous dysplasia, hereditary keratomas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea

35 and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, sensorineural hearing loss, and disorders of immune cell activation; a cell signaling disorder including

- endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalcemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, teratogenesis, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and fibrocystic breast disease; and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5  $\alpha$ -reductase, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.
- In another embodiment, a vector capable of expressing CCYPR or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR including, but not limited to, those described above.
- In a further embodiment, a pharmaceutical composition comprising a substantially purified CCYPR in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CCYPR may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CCYPR including, but not limited to, those listed above.

In a further embodiment, an antagonist of CCYPR may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CCYPR. Examples of such disorders include, but are not limited to, those immune, developmental, and cell signaling disorders, and cell proliferative disorders including cancer, described above. In one aspect, an antibody which specifically binds CCYPR may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CCYPR.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CCYPR may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CCYPR including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate

therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CCYPR may be produced using methods which are generally known in the art. In particular, purified CCYPR may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CCYPR. Antibodies to CCYPR may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CCYPR or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lyssolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (*Bacilli Calmette-Guerin*) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to



CCYPR have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of CCYPR amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CCYPR may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cole, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.)

Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CCYPR-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for CCYPR may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CCYPR and its

specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CCYPR epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CCYPR. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of CCYPR-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CCYPR epitopes, represents the average affinity, or avidity, of the antibodies for CCYPR. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular CCYPR epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the CCYPR-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunoprecipitation and similar procedures which ultimately require dissociation of CCYPR, preferably in active form, from the antibody (Catty, D.

(1988) *Antibodies, Volume I: A Practical Approach*, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) *A Practical Guide to Monoclonal Antibodies*, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CCYPR-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Colligan et al., *supra*.)

In another embodiment of the invention, the polynucleotides encoding CCYPR, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CCYPR. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CCYPR. (See, e.g., Agrawal, S., ed. (1996) *Antisense Therapeutics*, Humana Press Inc., Totowa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence

complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Ucker, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

10 In another embodiment of the invention, polynucleotides encoding CCYPR may be used for somatic or germ-line gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaise, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and Somia, N. (1997) *Nature* 389:239-242), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399). In the *brasilienis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in CCYPR expression or regulation causes disease, the expression of CCYPR from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

30 In a further embodiment of the invention, diseases or disorders caused by deficiencies in CCYPR are treated by constructing mammalian expression vectors encoding CCYPR and introducing these vectors by mechanical means into CCYPR-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vivo* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.*

9:445-450).

Expression vectors that may be effective for the expression of CCYPR include, but are not

limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF,

PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CCYPR may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus

(RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the

tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA

89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Biau (1998)

Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the

ecdysone-inducible promoter (available in the plasmids PVGXR and PIND; Invitrogen); the

FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V.

and H.M. Biau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous

gene encoding CCYPR from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID

TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver

polynucleotides to target cells in culture and require minimal effort to optimize experimental

parameters. In the alternative, transfection is performed using the calcium phosphate method

(Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al.

(1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these

standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with

respect to CCYPR expression are treated by constructing a retrovirus vector consisting of (i) the

polynucleotide encoding CCYPR under the control of an independent promoter or the retrovirus long

terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive

element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences

required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are

commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc.

Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an

appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for

receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.

(1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and

A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R.

et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Riggs ("Method for obtaining

retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CCYPR to cells which have one or more genetic abnormalities with respect to the expression of CCYPR. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csicsi, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinuzzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CCYPR to target cells which have one or more genetic abnormalities with respect to the expression of CCYPR. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CCYPR to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Lin, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to Deluca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of

herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to

deliver polynucleotides encoding CCYPR to target cells. The biology of the prototypic alphavirus,

Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotech.* 9:464-469). During alphavirus

RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the

overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease

and polymerase). Similarly, inserting the coding sequence for CCYPR into the alphavirus genome in

place of the capsid-coding region results in the production of a large number of CCYPR-coding RNAs and the synthesis of high levels of CCYPR in vector transduced cells. While alphavirus infection is

typically associated with cell lysis within a few days, the ability to establish a persistent infection in

hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic

replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S. A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of

CCYPR into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones

of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus

infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10

and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can

be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes

inhibition of the ability of the double helix to open sufficiently for the binding of polymerases,

transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have

been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.L. Carr,

Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A

complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of

RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme

molecule to complementary target RNA, followed by endonucleolytic cleavage. For example,

engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze

endonucleolytic cleavage of sequences encoding CCYPR.

- Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.
- Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding CCYPR. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.
- RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases. An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding CCYPR.
- Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CCYPR expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CCYPR may be therapeutically useful, and in the treatment of disorders associated with decreased CCYPR expression or activity, a compound which specifically promotes expression of the polynucleotide encoding CCYPR may be therapeutically useful.
- At least one, and up to a plurality, of test compounds may be screened for effectiveness in

- altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CCYPR is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding CCYPR are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CCYPR. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).
- Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)
- Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys. An additional embodiment of the invention relates to the administration of a pharmaceutical composition which generally comprises an active ingredient formulated with a pharmaceutically



acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Mack Publishing, Easton PA). Such pharmaceutical compositions may consist of CCYP<sub>R</sub>, antibodies to CCYP<sub>R</sub>, and mimetics, agonists, or inhibitors of CCYP<sub>R</sub>.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Pharmaceutical compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g.,

Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of pharmaceutical compositions may be prepared for direct intracellular delivery of macromolecules comprising CCYP<sub>R</sub> or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CCYP<sub>R</sub> or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) *Science* 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CCYP<sub>R</sub> or fragments thereof, antibodies of CCYP<sub>R</sub>, and agonists, antagonists or inhibitors of CCYP<sub>R</sub>, which

ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind CCYPR may be used for the diagnosis of disorders characterized by expression of CCYPR, or in assays to monitor patients being treated with CCYPR or agonists, antagonists, or inhibitors of CCYPR. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CCYPR include methods which utilize the antibody and a label to detect CCYPR in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CCYPR, including ELISAs, RIAs, and FACS, are known

in the art and provide a basis for diagnosing altered or abnormal levels of CCYPR expression. Normal or standard values for CCYPR expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to CCYPR under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CCYPR expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. In another embodiment of the invention, the polynucleotides encoding CCYPR may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CCYPR may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CCYPR, and to monitor regulation of CCYPR levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CCYPR or closely related molecules may be used to identify nucleic acid sequences which encode CCYPR. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CCYPR, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CCYPR encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:55-108 or from genomic sequences including promoters, enhancers, and introns of the CCYPR gene.

Means for producing specific hybridization probes for DNAs encoding CCYPR include the cloning of polynucleotide sequences encoding CCYPR or CCYPR derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CCYPR may be used for the diagnosis of disorders associated with expression of CCYPR. Examples of such disorders include, but are not limited to, an immune disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome

- (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asplenia, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholelithiasis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytopenia, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, mixed connective tissue disorder (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocytopenia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidosis, hereditary keratinodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, sensorineural hearing loss, and disorders of immune cell activation; a cell signaling disorder including endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalcemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, perturbations of the menstrual cycle, polycystic

ovarian disease, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, teratogenesis, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and fibrocystic breast disease; and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5  $\alpha$ -reductase, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynaecomastia; and a cell proliferative disorder such as

actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis,

thymus, thyroid, and uterus. The polynucleotide sequences encoding CCYPR may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CCYPR expression. Such qualitative or quantitative methods are well known in the art. In a particular aspect, the nucleotide sequences encoding CCYPR may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide

sequences encoding CCYPR may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CCYPR in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CCYPR, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CCYPR, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified

polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated,

5 hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

10 With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual

clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further

progression of the cancer.

15 Additional diagnostic uses for oligonucleotides designed from the sequences encoding CCYPR

may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding

CCYPR, or a fragment of a polynucleotide complementary to the polynucleotide encoding CCYPR, and will be employed under optimized conditions for identification of a specific gene or condition.

20 Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding CCYPR may be used to detect single nucleotide polymorphisms (SNPs). SNPs are

25 substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation

polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding CCYPR are used to amplify DNA using the

polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary

30 and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are

fluorescently labeled, which allows detection of the amplicons in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (iSSNP), are capable of identifying polymorphisms by comparing the sequence of individual

overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of CYPs include radiolabeling or biotinylation nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large

numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the

activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile. In another embodiment, antibodies specific for CYPs, or CYPs or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of

transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed

molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000)

*Toxicol. Lett.* 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data

after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present



invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

20 A proteomic profile may also be generated using antibodies specific for CCYPs to quantify the levels of CCYP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoz, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

25 Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Selthamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention. In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding CCYPR may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map

data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CCYP<sub>R</sub> on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

5 In situ hybridization of chromosomal preparations and physical mapping techniques, such as

linkage analysis using established chromosomal markers, may be used for extending genetic maps.

Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is

valuable to investigators searching for disease genes using positional cloning or other gene discovery

10 techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized

by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences

mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g.,

Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may

also be used to detect differences in the chromosomal location due to translocation, inversion, etc.,

15 among normal, carrier, or affected individuals.

In another embodiment of the invention, CCYP<sub>R</sub>, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug

screening techniques. The fragment employed in such screening may be free in solution, affixed to a

solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes

20 between CCYP<sub>R</sub> and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds

having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT

application WO84/03564.) In this method, large numbers of different small test compounds are

25 synthesized on a solid substrate. The test compounds are reacted with CCYP<sub>R</sub>, or fragments thereof,

and washed. Bound CCYP<sub>R</sub> is then detected by methods well known in the art. Purified CCYP<sub>R</sub> can

also be coated directly onto plates for use in the aforementioned drug screening techniques.

Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a

solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing

30 antibodies capable of binding CCYP<sub>R</sub> specifically compete with a test compound for binding CCYP<sub>R</sub>.

In this manner, antibodies can be used to detect the presence of any peptide which shares one or more

antigenic determinants with CCYP<sub>R</sub>.

In additional embodiments, the nucleotide sequences which encode CCYP<sub>R</sub> may be used in any

molecular biology techniques that have yet to be developed, provided the new techniques rely on

properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/145,075, U.S. Ser. No. 60/153,129, and U.S. Ser. No. 60/164,647, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic

oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pCDNA2.1 plasmid

(Invitrogen, Carlsbad CA), or pINCY plasmid (Incye Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, GaitHERSBURG MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (LabSystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incye cDNA recovered in plasmids as described in Example II were sequenced as follows.

Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions,

references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation

using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length

polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SWISSProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a

probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:55-108. Fragments from about 20 to about 400 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

#### IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel, 1995, *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related

molecules in cDNA databases such as GenBank or LIFESSEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length(Seq. 1)}, \text{length(Seq. 2)} \}}$$

5

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

15

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding CCYPR occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table

20

3. V. Chromosomal Mapping of CCYPR Encoding Polynucleotides

30

The cDNA sequences which were used to assemble SEQ ID NO:55-108 were compared with sequences from the Incyte LIFESSEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:55-108 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available

from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Génethon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

The genetic map locations of SEQ ID NO:61, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:81, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:104, and SEQ ID NO:105 are described in The Invention as ranges, or intervals, or human chromosomes. More than one map location is reported for SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:90, and SEQ ID NO:100, indicating that previously mapped sequences having similarity, but not complete identity, to SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:90, and SEQ ID NO:100 were assembled into their respective clusters. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.)

The cM distances are based on genetic markers mapped by Génethon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

#### VI. Extension of CCYPR Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:55-108 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme



(Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Action MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (LabSystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence. The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1:

94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:55-108 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such

extension, and an appropriate genomic library.

## VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:55-108 are employed to screen cDNAs,

genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments.

Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size

exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon

membranes (Nytan Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16

hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate.

Hybridization patterns are visualized using autoradiography or an alternative imaging means and

compared.

## VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing

photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned

technologies should be uniform and solid with a non-porous surface (Schna (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure

analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a

substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be

produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schna, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol.

16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may

comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be

selected using software well known in the art such as LASERGENE software (DNASTAR). The array

elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

10 Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with

15 GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### 25 Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR

35 Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and

coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cyt3 and Cyt5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample

mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for

about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

#### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an

Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cyt3 and at 632 nm for excitation of Cyt5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide

containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.

Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477,

Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cyt3 and 650 nm for Cyt5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical cross-talk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

**IX. Complementary Polynucleotides**

Sequences complementary to the CCYPR-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CCYPR. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CCYPR. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CCYPR-encoding transcript.

## **X. Expression of CCYPR**

Expression and purification of CCYPR is achieved using bacterial or virus-based expression systems. For expression of CCYPR in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*lac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express CCYPR upon induction with isopropyl beta-D-

thiogalactopyranoside (IPTG). Expression of CCYPR in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CCYPR by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (SF9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CCYPR is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CCYPR, at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunofluorescence purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified CCYPR, obtained by these methods can be used directly in the assays shown in Examples XI and XV.

#### XI. Demonstration of CCYPR Activity

An assay for CCYPR activity measures cell proliferation as the amount of newly initiated DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding CCYPR is transfected into quiescent 3T3 cultured cells using methods well known in the art. The transiently transfected cells are then incubated in the presence of [<sup>3</sup>H]thymidine, a radioactive DNA precursor. Where applicable, varying amounts of CCYPR ligand are added to the transfected cells. Incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA and CCYPR activity.

#### XII. Functional Assays

CCYPR function is assessed by expressing the sequences encoding CCYPR at physiologically

elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CCYPR on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CCYPR and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CCYPR and other genes of interest can be analyzed by northern analysis or microarray techniques.

**XIII. Production of CCYPR Specific Antibodies**

CCYPR substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Hartington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CCYPR amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is

synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CCYPR activity by, for example, binding the peptide or CCYPR to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

**XIV. Purification of Naturally Occurring CCYPR Using Specific Antibodies**

Naturally occurring or recombinant CCYPR is substantially purified by immunoaffinity chromatography using antibodies specific for CCYPR. An immunoaffinity column is constructed by covalently coupling anti-CCYPR antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CCYPR are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CCYPR (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CCYPR binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CCYPR is collected.

#### **XV. Identification of Molecules Which Interact with CCYPR**

CCYPR, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CCYPR, washed, and any wells with labeled CCYPR complex are assayed. Data obtained using different concentrations of CCYPR are used to calculate values for the number, affinity, and association of CCYPR with the candidate molecules.

Alternatively, molecules interacting with CCYPR are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, *Nature* 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech). CCYPR may also be used in the PATHCALLING process (Curagen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent



No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. 5 Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	55	116462	KIDNNOT01	116462H1 (KIDNNOT01), 116462R1 (KIDNNOT01), 116462X304D1 (KIDNNOT01), 1500439F6 (SINTBST01), 2369977F6 (ADRENOT07)
2	56	1210462	BRSTNOT02	260707H1 (HNT2RAT01), 1210462H1 (BRSTNOT02), 1458882F6 (COLNFET02), 1841248T6 (COLNNOT07), 2378362H1 (ISLTNOT01), 3728643F6 (SMCCNON03)
3	57	1305252	PLACNOT02	794067R6 (OVARNOT03), 871989R1 (LUNGAST01), 1235253F1 (LUNGFET03), 1305252F6 (PLACNOT02), 1305252H1 (PLACNOT02), 1703258T6.comp (DUODNOT02), 2678307H1.comp (OVARUT07), 3221088H1.comp (COLNNOT03), 3647280H1 (ENDINOT01)
4	58	1416289	BRAINOT12	639958R6 (BRSTNOT03), 861752H1 (BRAITUT03), 1416289H1 (BRAINOT12), 1416289X310B1 (BRAINOT12), 1416289X310D2 (BRAINOT12), 1947451R6 (PITVNOT01)
5	59	1558289	SPLNNOT04	1558289H1 (SPLNNOT04), 1852450T6 (LUNGFET03), 2396092F6 (THP1AZT01), 2593267F6 (LUNGNOT22), 2632784F6 (COLNTUT15)
6	60	1577739	LNODNOT03	181266R1 (PLACNOB01), 1577739H1 (LNODNOT03), 4180022T6 (SINITUT03), 4597046H1 (COLSTUT01), 4860616H1 (PROSTUT09), 4991290H1 (LIVRTUT11), 5059810H1 (CONDTUT02)
7	61	1752768	LIVRTUT01	256106R1 (HNT2RAT01), 258814H1 (HNT2RAT01), 1312247F1 (COLNFET02), 1344279T6 (PROSNOT11), 1350089H1 (LATRTUT02), 1440718F6 (THYRNOT03), 1752768F6 (LIVRTUT01), 1752768H1 (LIVRTUT01), 1752768T6 (LIVRTUT01), 2079106F6 (ISLTNOT01), SBYA00612U1
8	62	1887228	BLADTUT07	080294F1 (SYNOBAB01), 140055F1 (TLYMNOR01), 285207X42 (EOSIHET02), 516882R6 (MMLR1DT01), 1217892T1 (NEUTGMT01), 1887228H1 (BLADTUT07), 4323029H1 (TLYMUNT01)
9	63	1988468	LUNGAST01	072147R6 (THP1PEB01), 496297H1 (HNT2NOT01), 1362109F6 (LUNGNOT12), 1726095F6 (PROSNOT14), 1726095T6 (PROSNOT14), 1988468H1 (LUNGAST01), 1988468T6 (LUNGAST01), 2232471F6 (PROSNOT16)
10	64	2049176	LIVRFET02	2049176H1 (LIVRFET02), 2049176T6 (LIVRFET02), 2049176X321D1 (LIVRFET02)
11	65	2686765	LUNGNOT23	1502858F6 (BRAITUT07), 1956694X315D1 (CONNNOT01), 2022628X307D1 (CONNNOT01), 2686765F6 (LUNGNOT23), 2686765H1 (LUNGNOT23), 2864555H1 (KIDNNOT20), 2887609F6 (SINJNOT02), 3381980H1 (ESOGNOT04)
12	66	3215187	TESTNOT07	151135R6 (FIBRAGT01), 3215187F6 (TESTNOT07), 3215187H1 (TESTNOT07)
13	67	3500375	PROSTUT13	860585R1 (BRAITUT03), 1318501F1 (BLADNOT04), 1419126F1 (KIDNNOT09), 1483246F6 (CORPNOT02), 2238114T6 (PANCYTUT02), 2272329H1 (PROSONOT1), 3209746F7 (BLADNOT08), 3403213H1 (ESOGNOT03), 4176619H1 (BRAINOT22), 4614606H1 (BRAYDTUT01)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
14	68	5080410	LNODNOT11	1270372X300D1 (BRAINOT09), 3460603H1 (293TF1T01), 5080410H1 (LNODNOT11)
15	69	5218248	BRSTNOT35	1808748X15C1 (PROSTUT12), 1808748X16C1 (PROSTUT12), 3391884H1 (LUNGNOT28)
16	70	058336	MUSCNOT01	058336H1 (MUSCNOT01), 058336T6 (MUSCNOT01), g2206766, g2069225
17	71	1511488	LUNGNOT14	1436265F1 (PANCNOT08), 1511488H1 (LUNGNOT14), 1511488T6 (LUNGNOT14), 1850020F6 (LUNGFEET03)
18	72	1638819	UTRSNOT06	1282638T1 (COLNNOT16), 1638819F6 (UTRSNOT06), 1638819H1 (UTRSNOT06), 3597071H1 (FIBPNOT01), SBRA03813D1, SBRA04133D1, SBRA03785D1
19	73	1655123	PROSTUT08	1271351F1 (TESTTUT02), 1353234F1 (LATRTUT02), 1655123H1 (PROSTUT08), 2132186R6 (OVARNOT03), 3296525H1 (TLXJINT01), 3354010H1 (PROSNOT28), 3741838F6 (MENTNOT01), 3741838T6 (MENTNOT01), SXAF03528V1
20	74	2553926	THYMNOT03	403261F1 (TMLR3D1T01), 1869739F6 (SKINBIT01), 2197242T6 (SPLNFEET02), 2553926H1 (THYMNOT03), 2553956T6 (THYMNOT03), 3935528H1 (PROSTUT09), 5263918F6 (CONDTUT02)
21	75	2800717	PENCNOT01	411179F1 (BRSTNOT01), 415284R1 (BRSTNOT01), 1458971F1 (COLNFEET02), 1600810H1 (BLADNOT03), 1622005F6 (BRAITUT13), 2173076F6 (ENDCNOT03), 2520087F6 (BRAITUT21), 2800717H1 (PENCNOT01), 5184583H1 (LUNGMT03), 5435834H1 (SPLNNOT17), 5872662H1 (COLTDIT04)
22	76	5664154	BRAUNOT01	181534F1 (PLACNOB01), SGHA00262V1
23	77	017900	HUVELPB01	017900H1 (HUVELPB01), 092858F1 (HYPONOB01), 1353543F1 (LATRTUT02), 1353543F6 (LATRTUT02), 1428464F1 (SINTBST01), g1616429
24	78	035102	HUVENOB01	035102H1 (HUVENOB01), 077722R1 (SYNORAB01), 995133H1 (KIDNTUT01), 1356968T6 (LUNGNOT09), 1963135R6 (BRSTNOT04), 2659921F6 (LUNGUT09), 3110603H1 (BRSTNOT17)
25	79	259983	HNT2RAT01	259131R1 (HNT2RAT01), 259983H1 (HNT2RAT01), 268205R1 (HNT2NOT01), 1305726F1 (PLACNOT02)
26	80	926810	BRAINOT04	926810H1 (BRAINOT04), 3490378T6 (EPIGNOT01), 4774848H1 (BRAQNOT01), SBIA01080D1, SBIA04006D1, SBIA02273D1, SBIA01121D1
27	81	1398816	BRAITUT08	056398F1 (FIBRNOT01), 1252138F2 (LUNGFEET03), 1294556T1 (PGANNOT03), 1398816H1 (BRAITUT08), 1545328R1 (PROSTUT04)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
28	82	1496820	PROSNON01	996673H1 (KIDNTUT01), 1496820H1 (PROSNON01), 2368484F6 (ADRENOT07), 3071781X303D1 (UTRSNOR01), 3071781X307B1 (UTRSNOR01), 3071781X316B2 (UTRSNOR01), 3071781X316D3 (UTRSNOR01)
29	83	1514559	PANCTUT01	155768H1 (THP1PLB02), 1229952H1 (BRAITUT01), 1337018X11 (COLNNOT13), 1360361H1 (LUNGNOT12), 1365811H1 (SCORONOT2), 1514559F6 (PANCTUT01), 1514559H1 (PANCTUT01)
30	84	1620092	BRAITUT13	1620092F6 (BRAITUT13), 1620092H1 (BRAITUT13), 1832842H1 (BRAINON01), 1843815R6 (COLNNOT08), 1843815T6 (COLNNOT08)
31	85	1678765	STOMFET01	1678765F6 (STOMFET01), 1678765H1 (STOMFET01), 2640786H1 (LUNGUTUT08), 3542276F6 (TONSNOT03), 4180591H1 (SINITUT03), 4183383H1 (LIVRDIR01), 4349212H1 (TLYMTXT01), 4718559H1 (BRAIHCT02), 5023762H1 (OVARNON03), 5332272H1 (KIDNNOT34), 91665766
32	86	1708229	PROSNOT16	388493R1 (THYMNOT02), 1503519F1 (BRAITUT07), 1708229H1 (PROSNOT16), 1725267F6 (PROSNOT14), 3089258F6 (HEONOT03)
33	87	1806454	SINTNOT13	406723H1 (EOSIHET02), 821556R1 (KERANOT02), 1649621F6 (PROSTUT09), 1710552H1 (PROSNOT16), 1806454F6 (SINTNOT13), 1806454H1 (SINTNOT13), 2526283H1 (BRAITUT21), 3869969H1 (BMAARNOT03)
34	88	1806850	SINTNOT13	270548H1 (HNT2NOT01), 443885R1 (MPHGNOT03), 1257235F1 (MENITUT03), 1337438H1 (COLNNOT13), 1351820F1 (LATRTUT02), 1544066R1 (PROSTUT04), 1806850F6 (SINTNOT13), 1806850H1 (SINTNOT13), 1984108T6 (LUNGAST01), 2921419H1 (SININOT04), 3109392H1 (BRSTUT15)
35	89	1851534	LUNGFET03	1851534H1 (LUNGFET03), 2407346R6 (BSTMNON02), 2757389R6 (THP1AZS08), 5513454H1 (BRADIR01), 5629312H1 (PLACFER01)
36	90	1868749	SKINBIT01	1322048F1 (BLADNOT04), 1398330F1 (BRAITUT08), 1437866F6 (PANCNOT08), 1868749F6 (SKINBIT01), 1868749H1 (SKINBIT01), 2279968R6 (PROSNON01), 2684670H1 (LUNGNOT23), 4632232H1 (GBLADIT02), 4951533H2 (ENDVUNT01), 5077673H1 (LNODNOT11), 5388496H1 (BRAINOT19)
37	91	1980010	LUNGUTUT03	127747R1 (TESTNOT01), 357561F1 (PROSNOT01), 357561R1 (PROSNOT01), 918017R1 (BRSTNOT04), 142817F6 (SINTBST01), 1625080F6 (COLNPOT01), 1720753H1 (BLADNOT06), 1932038F6 (COLNNOT16), 1980010H1 (LUNGUTUT03), 3112417F6 (BRSTNOT17), 4174704H1 (SINTNOT21), 4238802H1 (SYNWDIT01), 5499543H1 (BRADIR01), 94337459

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
38	92	2259032	OVARTUT01	475134H1 (MLLR2DT01), 784284R1 (PROSNOT05), 1264124H1 (SYNORAT05), 1418710F1 (KIDNNOT09), 1697570T6 (BLADTUT05), 1874051F6 (LEUKNOT02), 2187960T6 (PROSNOT26), 2259032H1 (OVARTUT01), 2259032R6 (OVARTUT01), 3406237H1 (ESOGNOT03), 3441729H1 (PENCNOT06), 355764H1 (LUNGNOT31), 3728010H1 (SMCCNOT03), 3813639H1 (TONSNOT03), 4031501H1 (BRAINOT23), 4274704H1 (PROSTMT01), 4602450H1 (BRSTNOT07), 93327183
39	93	2359526	LUNGFET05	1667182F6 (BMARNOT03), 2359526H1 (LUNGFET05), 2359526X311D1 (LUNGFET05), 2555305F7 (THYMNOT03), 2654667T6 (THYMNOT04), SCHA00290V1, SCHA00266V1, 91748241
40	94	2456494	ENDANOT01	1860223F6 (PROSNOT18), 2456494H1 (ENDANOT01), 2564671H1 (ADRETUT01), 3618339H1 (EPIPNOT01)
41	95	2668536	ESOGTUT02	1513847H1 (PANCUTUT01), 1668943F6 (BMARNOT03), 1668943T6 (BMARNOT03), 1721443F6 (BLADNOT06), 2668536H1 (ESOGTUT02), 3438287H1 (PENCNOT05), SBFA00330F1, SCBA05255V1, SCBA01530V1
42	96	2683225	SINIUCT01	196443R6 (KIDNNOT02), 1243440R6 (LUNGNOT03), 1604540F6 (LUNGNOT15), 2072837H1 (ISLTNOT01), 2683225F6 (SINIUCT01), 2683225H1 (SINIUCT01), 3647874H1 (ENDINOT01), 4029178H1 (BRAINOT23)
43	97	2797839	NPOLNOT01	460779T6 (KERANOT01), 782663H1 (MYONOT01), 896898R1 (BRSTNOT05), 1218533H1 (NEUTGMT01), 1312923F6 (BLADTUT02), 2473746F6 (THP1NOT03), 2481564H1 (SMCANOT01), 2797839H1 (NPOLNOT01), 3350118H1 (BRAITUT24), 4184264H1 (BRABDIR01), 4401265H1 (TESTTUT03), 4727770H1 (GBLADIT01), 5080203H1 (LNODNOT11), 5524886H1 (LIVRDIR01)
44	98	2959521	ADRENOT09	046696H1 (CORNNOT01), 087727R6 (LIVRNOT01), 138475H1 (LIVRNOT01), 167505H1 (LIVRNOT01), 647975H1 (CARCTXT02), 781084T1 (MYOMNOT01), 972191R6 (MUSCNOT02), 1309196H1 (COLNFET02), 2641117H1 (LUNGUTUT08), 2913953H1 (KIDNUTUT15), 2959521H1 (ADRENOT09), 2984654H1 (CARGDIT01), 2985141H1 (CARGDIT01), 3138371H1 (SMCCNOT02), 3386016H1 (ESOGNOT04), 3496187H1 (ADRETUT07), 3614426H1 (EPIPNOT01), 4287819H1 (LIVRDIR01), 5395566H1 (LIVRUT13), 9505101
45	99	3082014	BRAIUNT01	182588H1 (PLACNOB01), 645276R6 (BRSTTUT02), 1497811F1 (SINTBST01), 2051505F6 (LIVRFET02), 3082014H1 (BRAIUNT01), 3464112F6 (293TF2T01), 4603079H1 (BRSTNOT07)
46	100	3520701	LUNGNO3	971201H1 (MUSCNOT02), 1544657R6 (PROSTUT04), 1545570H1 (PROSTUT04), 1671030F6 (BMARNOT03), 1671030T6 (BMARNOT03), 2605263F6 (LUNGUTUT07), 3520701H1 (LUNGNO3), 3520701R6 (LUNGNO3)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
47	101	4184320	BRADDIT02	215656F6 (BRAINOT09), 4184253F6 (BRABDIR01), 4184253T6 (BRABDIR01), 4184320H1 (BRADDIT02), 4252542F6 (BRABDIR01)
48	102	4764233	PLACNOT05	4764233H1 (PLACNOT05), 5634642H1 (PLACFER01), 91148809
49	103	4817352	HELATXT03	426993R6 (BLADNOT01), 426993T6 (BLADNOT01), 488301R6 (HNTAGT01), 3779640H1 (BRSTNOT27), 4817352H1 (HELATXT03)
50	104	5040573	COLHTUT01	1724126F6 (PROSNOT14), 1859337F6 (PROSNOT18), 2026289R6 (KERANOT02), 2026289T6 (KERANOT02), 2122846T6 (BRSTNOT07), 3225302H1 (ADRETUT07), 3322214H1 (PTHYNOT03), 4587178H1 (BRAQNOT01), 4601227H1 (BRSTNOT07), 4885408H1 (LUNLMT01), 5040573H1 (COLHTUT01)
51	105	5627029	PLACFER01	967988R1 (BRSTNOT05), 1534642T6 (SPLNNOT04), 1700904F6 (BLADTUT05), 1846971R6 (COLNNOT09), 2112727R6 (BRAITUT03), 2112727T6 (BRAITUT03), 2205225F6 (SPLNFET02), 2828475H1 (TLYNNOT03), 3439165F6 (PENCNOT06), 3604622H1 (LUNGNOT30)
52	106	5678487	293TF2T01	1258787F6 (MENITUT03), 1522008F1 (BLADTUT04), 1597992F6 (BLADNOT03), 2057679H1 (BEPINOT01), 2411504H1 (BSTMNON02), 2467956H1 (THYRNOT08), 2739089F6 (OVARNOT09), 2739089T6 (OVARNOT09), 2740762H1 (BRSTTUT14), 2754616H1 (THPIAZS08), 3254971R6 (OVARFUN01), 3487616H1 (EPIGNOT01), 5678487H1 (293TF2T01)
53	107	5682976	BRAENOT02	3504932H1 (LVEENNOT01), 825361R1 (PROSNOT06), 879866R1 (THYRNOT02), 1667502F6 (BMAARNOT03), 1733323F6 (BRSTTUT08), 1876248T6 (LEUKNOT02), 1963215T6 (BRSTNOT04), 2539188H1 (BONRTUT01), 2896448H1 (KIDNTUT14), 3141553H1 (SMCCNOT02), 3374826F6 (CONNTUT05), 3773427H1 (BRSTNOT25), 3779981H1 (BRSTNOT27), 5682976H1 (BRAENOT02), 5546853H1 (TESTNOC01)
54	108	5992432	FTUBTUT02	645878R6 (BRSTTUT02), 1287660F1 (BRAINOT11), 1287660T6 (BRAINOT11), 1417373F6 (BRAINOT12), 1618868F6 (BRAITUT12), 2269980R6 (UTRSNOT02), 2793117F6 (COLNTUT16), 3246793F6 (BRAINOT19), 3592787H1 (293TF5T01), 5992432H1 (FTUBTUT02), 9821012

Table 2

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
1	145	T10 S93	M15 N38	Signal peptide: M1-Q33 Protein SH3 domain repeat: L8-R99 GLGF signal transduction-related domain: M1-R99		MOTIFS SPSCAN BLAST_PRODOM BLAST_DOMO
2	340	T39 S190 S268 T307 S88 S102 S165 S226 S230 S234 T337		P120 nuclear proliferating cell antigen: N117-K333 Proliferative cell nucleolar protein P120: E26-G293	Proliferating cell nucleolar antigen P120 (g26649749) <u>A. fulgidus</u>	MOTIFS BLAST_PRODOM BLAST_DOMO BLAST_GenBank
3	418	S246 S415 T142 T156 S292 S349 S369 S64 S247 S298	N190 N191 N203 N288 N306		Candidate tumor suppressor p33ING1 (g2829208) <u>H. sapiens</u>	MOTIFS BLAST_GenBank
4	297	T217 T82 S76 S127 S176 T207 S246 Y189	N74	Germ cell-less protein: E96-N297	Germ cell-less protein (g5814404) <u>Mus musculus</u>	MOTIFS BLIMPS_PFAM BLAST_GenBank
5	184	T34 S103 S5 T136	N76		Differentiation factor MDC-3.13 (g3860093) <u>H. sapiens</u>	MOTIFS BLAST_GenBank
6	173	S109 S24 S59 S66 S141 S142 T152			Posterior end mark-5 (g4107015) <u>C. savignyi</u>	MOTIFS BLAST_GenBank

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
7	591	S582 T71 T208 S217 S339 T475 S493 T536 S45 S105 S153 T208 S305 S336 T578 Y93	N374 N425 N534 N585	Signal peptide M1-L64 TPR domain mitosis control E239-P356	Cell division cycle protein 23 homolog (g5541721) <u>A. thaliana</u>	MOTIFS SPSCAN HMMR_Pfam BLAST_DOMO BLAST_GenBank
8	463	T237 S34 T67 T117 T125 S138 T288 T321 S328 S418 T80 S186 S190 S209 S210 T232 T288 S418 T441 S445 Y416	N208	TPR repeat V265-K516 Formin limb deformity: M1-E335	Lymphocyte specific formin related protein (g4101720) <u>M. musculus</u>	MOTIFS BLAST_PRODUM BLAST_DOMO BLAST_GenBank
9	270		N64 N94 N147		Early embryogenesis MRG1 protein (g2570051) <u>M. musculus</u>	MOTIFS BLAST_GenBank
10	255	S180 T49 T53 S97 S152 T201 S210 S23 S97 T145 T216 S225 S228 T231 S242 Y106 Y240		Polypopsis locus TB2 homolog: G15-T117 Polypopsis locus protein: V13-T117	Similar to polypopsis locus protein 1 (g849238) <u>H. sapiens</u>	MOTIFS BLAST_PRODUM BLAST_DOMO BLAST_GenBank
11	533	S227 S412 S505 S7 S17 S65 T349 S442 T29 S72 S89 S358 S442 T446 S505 Y244		TRE oncogene: R56-I277	TRE oncogene-related protein (g2286196) <u>D. melanogaster</u>	MOTIFS BLOCKS_DOMO BLAST_GenBank



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
12	160	S40		Signal peptide: M1-A30 Transmembrane domain: A6-I29 Cornichon developmental protein: M1-S160	Cornichon-like protein (g4521254) <u>M. musculus</u>	MOTIFS SPSCAN HMMR BLAST_PRODOM BLAST_DOMO BLAST_GenBank
13	531	S195 T196 S357 T45 S172 T199 S212 S322 S465 T495 T45 T241 S255 T279 T319 S328	N244 N401		Cdc 73p (g632679) <u>S. cerevisiae</u>	MOTIFS BLAST_GenBank
14	165	S3 T67 S104			Wolf-Hirschhorn syndrome candidate 2 protein (g3860187) <u>H. sapiens</u>	MOTIFS BLAST_GenBank
15	199	S2 S21 S69 T102 S189			Developmental protein DG1118 (g3789911) <u>D. discoideum</u>	MOTIFS BLAST_GenBank
16	168	S141 S55 S61 T79	N77	Signal peptide M1-S61 H-Rev protein homolog P15-K166	g3777529 retinoic acid receptor responder 3 <u>Homo sapiens</u>	BLAST_GenBank SPSCAN BLAST_PRODOM MOTIFS
17	162	S70 S85 T16 T28 T65 T80 T100 S127 Y111			g207250 growth and transformation dependent protein <u>Rattus norvegicus</u>	BLAST_GenBank

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
18	246	T209 S227 T243 T28 S223 S51 S136 S201	N26 N158	Protein cell intergenic region FTSJ K25-K241	g2622903 cell division protein J <u>Methanobacterium thermoauto-trophicum</u> g1322234 OS-9 precursor <u>Homo sapiens</u>	BLAST-GenBank BLAST-PRODOM BLAST-DMO MOTIFS
19	483	T394 T85 S86 S219 S225 T230 S298 T299 T472 S114 S200 T273 S371 T407 T424 T431		Signal peptide M1-G29 OS-9 precursor L54-E281	g3901272 ZW10 Interactor Zwint <u>Homo sapiens</u>	BLAST-GenBank SPSCAN BLAST-PRODOM MOTIFS
20	280	T129 T6 T102 T119 T181 S250 S46 T72 T84 S262		Signal peptide M1-L28	g3901272 ZW10 Interactor Zwint <u>Homo sapiens</u>	BLAST-GenBank SPSCAN MOTIFS
21	425	S122 S235 T60 S192 S203 S204 S218 S226 S307 T313 S332 S366 S370 T375 T402 S409 S89 S118 S241 S284 T360 Y399	N190 N311		g455719 Activated c-raf oncogenic fusion protein homolog <u>Homo sapiens</u>	BLAST-GenBank
22	128	S3 S107	N42	Prenyl group binding site (CAAX box) C125-P128 Ovarian granulosa cell 13.0 KD protein HGR74 N16-P128	g4580592 brain expressed X-linked protein 2 <u>Mus musculus</u>	BLAST-GenBank MOTIFS BLAST-PRODOM
23	113	S88 T20 T37		Biotin-requiring enzyme attachment site: L40-L90	LD0C-1 protein g3869127 ( <u>Homo sapiens</u> ) Nagasaki, K. et al. (1999) Cancer Lett. 140:227-234.	BLAST-GenBank PROFILESCAN MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
24	308	S95 T79 T98 S184 S246 S251 T55 S184 S226 S294 S300 Y127	N77	Melanoma antigen gene (MAGE) family: M1-Q200, H205-D283, D91-A287	Breast cancer associated gene 1 g4928044 (Homo sapiens) Lurquin, C. et al. (1997) Genomics 46:397-408.	BLAST-GenBank BLAST-PRODOM HMMER-PFAM BLAST-DOMO MOTIFS
25	221	S145 S160 S217 S25 S31 S70 S85 T89 S153 S197 Y34	N139	Annexin VI signature: L86-V95 Sushi domain: T165-C174	Teratocarcinoma expressed gene Tera g1575505 (Mus musculus)	BLAST-GenBank BLIMPS-PRINTS BLIMPS-PFAM MOTIFS
26	402	T344 S39 S78 S109 S237 T269 S273 T376 T381 T383 S11 S49 T89 T344 S364	N76 N107 N171 N362		Paraneoplastic cancer-testis-brain antigen g6179740 (Homo sapiens)	BLAST-GenBank MOTIFS
27	93	S11			Hypoxia inducible gene-1 g4929330 (Homo sapiens)	BLAST-GenBank MOTIFS
28	353	S125 T42 S43 S85 S212 S283 S314 T42 S49 S105 S120 S133 S162 S163 S212 S290	N145 N157 N191	af-4 (FEL protein): S195-K353 E4-Q185	AF5q31 protein g6601438 (Homo sapiens)	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS
29	120	T57		Cyclin-dependent kinase inhibitor: D7-P106, M1-N114	Cyclin dependent kinase inhibitor CIP1 g2276312 (Homo sapiens)	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS

Table 2 (cont.)

Polypeptide ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
30	144	S15 S64		Transmembrane domain: I93-I110	Transformation dependent protein g207250 (Rattus norvegicus) N.Glaichenhaus and F.Cuzin (1987) Cell 50:1081-1089.	BLAST-GenBank MOTIFS HMMER
31	933	S603 T51 S109 T129 S162 S203 S223 S224 S240 S261 S266 S280 S282 S313 T328 S346 S353 S378 S394 S460 S491 S499 T531 S627 S641 S642 S725 T732 S759 S188 S309 S423 S592 S671 S675 T706 S771 Y856	N107 N238 N639 N883		Replication protein Smp2 g218488 (Saccharomyces cerevisiae) Irie, K. et al. (1993) Mol. Gen. Genet. 6:283-288.	BLAST-GenBank MOTIFS
32	268	S7 T104 T154 S169	N90	Serine-Threonine kinase Binder MP51: L74-I230	Putative mitotic protein (Schizosaccharomyc es pombe) g3947877 F.C. Luca and M.Winey (1998) Mol Biol Cell 9:29-46.	BLAST-GenBank BLAST-PRODOM MOTIFS

Table 2 (cont.)

Polypeptide Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
33	337	T29 S236 T44 T238		Leucine zipper: L259-L280, L266-L287	DNA binding protein g184390 (Homo sapiens) Weitzel, J.N. et al. (1992) Genomics 14:309-319.	BLAST-GenBank MOTIFS
34	565	T17 S34 S61 S66 T138 T142 S174 T238 S245 S265 S436 S466 S527 S106 S205 S218 S258 T297 S314 T325 S463 T470 Y460	N347 N386 N506	F-Box domain: H75-Y123, L82-N95 Disease resistance protein: G254-I270	F-box protein FLR1 g7672734 (Homo sapiens)	BLAST-GenBank HMMER_Pfam BLIMPS-PRINTS MOTIFS
35	228	S200 T47 T62 S78 S107 S188 S192 S206 S200 S205 S213	N36 N94 N225		Predicted WHSC1 protein (Wolf-Hirschhorn syndrome critical region 1) g4378022 (Homo sapiens) Stecc I. et al. (1998) Hum. Mol. Genet. 7:1071-1082.	BLAST-GenBank MOTIFS
36	495	S451 S152 S365 S478 S108 S171 S181 T192 T347 T409 S435 Y86 Y111 Y203			Malignant brain tumor protein 1(3)mbt g3811111 (Homo sapiens) Koga, H. et al. (1999) Oncogene 18:3799-3809.	BLAST-GenBank MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
37	1336	T635 T769 S902 S10 S32 S33 T76 S95 S156 T298 S313 T427 S467 T579 T626 T642 S661 T668 S680 T699 T729 S774 S834 T859 T915 S944 S959 S961 S997 S1049 T1085 S1132 S1227 T1245 S1249 T48 S94 T169 S224 T352 T379 T389 T475 T696 S867 T883 T889 S940 S961 S1220 Y631	N148 N152 N345 N385 N1213 N1247	Ribosomal protein S14 signature: R1172-N1194 Leucine zipper: L211-L232	Neuroblastoma related protein g4337460 (Homo sapiens)	BLAST-GenBank BLIMPS-PRINTS MOTIFS
38	934	T532 S11 T23 T80 S171 S202 T214 T240 S244 T275 S412 S416 S437 S518 T523 S719 S746 S753 S796 S807 S93 T279 T527 S598 T780	N8 N210 N426	SAP: I92-Q364	Sap2 family putative cell cycle dependent phosphatase g3426127 (Schizosaccharomyces pombe) Luke, M.M. et al. (1996) Mol. Cell Biol. 16:2744-2755.	BLAST-GenBank BLAST-DOMO MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
39	515	T72 S122 S175 S272 S277 S305 T420 S422 T432 T79 S139 T189 S215 T316 S457 T486 Y13 Y383	N16 N31 N115	Metastasis-Associated Protein: E65-R230 Leucine zipper: L234-L255	Metastasis associated gene g1008544 (Homo sapiens) Toh, Y. et al. (1995) Gene 159:97-104 Toh, Y, et al. (1994) J Biol. Chem. 269:22958-22963.	BLAST-GenBank BLAST-PRODOM BLIMPS-PRINTS MOTIFS
40	146	S61		Leucine zipper: L5-L26, L12-L33, L19-L40	LD0C1 g3869127 (Homo sapiens)	BLAST-GenBank BLIMPS-PFAM MOTIFS
41	580	S324 S36 S340 S550 S86 T109 T119 T150 T226 S329 S340	N190	Cyclin: H19-K262	Cyclin K g3746549 (Homo sapiens) Edwards, M.C. et al. (1998) Mol. Cell Biol. 18:4291-4300.	BLAST-GenBank BLAST-PRODOM MOTIFS
42	131	S78 T121 T26		Presenilin: Q64-K75	Cell growth regulator DRR1 g4322559 (Homo sapiens) G.Thomas and M.N.Hall (1997) Curr. Opin. Cell Biol. 9:782-787.	BLAST-GenBank BLIMPS-PRINTS MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
43	812	S44 S588 S646 S801 S111 S120 S134 T140 S148 S150 S181 T185 S262 S279 S440 T477 S497 T520 T542 T605 S675 S40 T64 T311 T316 T319 T505 S562 S565 T566 T695 S702 S707 S708 T739 T776 S790 Y277	N503 N618	NOL1/NOP2/Fnu(sun) family signature: F454-G467, F300-K585, I388-M402, G410-G433, F454-G467, K507-L532, E189-M576 Proliferating Cell Nucleolar Antigen P120: M1-S134, E135-T311, F587-G805	Proliferating cell nuclear protein P120 g287723 (Homo sapiens)	BLAST-GenBank BLAST-PRODOM BLAST-DOMO BLIMPS-BLOCKS MOTIFS HMMER-PFAM
44	537	S505 T69 S138 S194 S310 S337 S356 T386 S485 S37 T45 T282	N122 N132 N147	Transmembrane domains: I506-G532, V271-L290, W472-F490	Estrogen induced protein in breast cancer LIV-1 g1256001 (Homo sapiens)	BLAST-GenBank HMMER MOTIFS
45	584	S185 T324 S343 T537 S575 S17 T102 S128 T229 T374 S412 T450	N28	Cytochrome C motif: C283-T288 Metastasis-associated protein MTA1: R19-R143, D144-K321, G340-G483, P432-K555 Leucine zipper: L147-L168	Metastasis associated gene g1008544 (Homo sapiens) Toh, Y. et al. (1995) Gene 159:97-104 Toh, Y. et al. (1994) J. Biol. Chem. 269:22958-22963.	BLAST-GenBank BLAST-PRODOM MOTIFS



Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
46	425	S190 T301 S12 S19 S41 S205 T206 T235 S263 S265 T315 S43 S52 S85 T93 T351 S411 Y422	N275	ML02 mitosis- associated protein: L24-R188, P226-Y245, N308-E408		BLAST-PRODOM MOTIFS
47	255	T9 T147 S237	N144	Melastatin: M1-R172, G199-G255	Melastatin g3047242 (Mus musculus) Duncan, L.M. et al. (1998) Cancer Res. 58:1515-1520.	BLAST-GenBank BLAST-PRODOM MOTIFS
48	111	T30 S2 T8			Melanoma associated antigen GAGE-8 g3511023 (Homo sapiens) Van den Eynde, B. et al. (1995) J. Exp. Med. 182:689-698.	BLAST-GenBank MOTIFS
49	422	T110 T159 S136 S150 T163 T190 S383 T413 S9 T27 S46 S96 T347 S359 S363 S368 Y350		XPMC2 (mitosis associated inducing protein): A236-E402	Mitotic regulator XPMC2 (Xenopus gene which prevents mitotic catastrophe) g595380 (Xenopus laevis) J.Y.Su and J.L.Maller (1995) Mol. Gen. Genet. 246:387-396.	BLAST-GenBank BLAST-PRODOM BLAST-DBOM MOTIFS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
50	397	S20 S21 T395 T57 S59 T64 S127 S208 T210 S262 S307 T341 T64 T168 S180 S185 S218 S231 S288 S326	N222 N260	Transmembrane motifs: I361-L380, L24-L44 Cell division control protein: K17-L347	Cell cycle protein CDC1 g550426 (Saccharomyces cerevisidae)	BLAST-GenBank HMMER BLAST-PRODOM MOTIFS
51	800	S56 S448 T721 S760 S48 S84 S111 S119 T146 T189 T235 S250 S265 T275 S321 S335 T392 S448 T466 S474 T562 S596 S598 T626 S686 S3 S4 S65 S89 S107 T123 S348 T398 T402 T716 S730 S738 T743 S789 Y102 Y316 Y569 Y685	N554 N665	Signal peptide: M1-A25 Leucine zipper: L365-L386	SART-1 g4126469 (Mus musculus)	BLAST-GenBank SPSCAN MOTIFS
52	713	S100 T631 S8 T9 S20 T42 T114 T121 T172 T177 T191 T192 S218 T231 T256 S325 S335 S381 T464 T482 T538 T581 T617 S693 S94 S166 T201 S202 S321 T568 S614 T658 Y459	N7 N49 N462	Leucine zipper: L680-L701	Colon cancer antigen NY-CO-8 g3170180 (Homo sapiens) Scanlan, M.J. et al. (1998) Int. J. Cancer 76:652-658.	BLAST-GenBank BLAST-PRODOM BLAST-DBOM MOTIFS

Table 2 (cont.)

Polypeptide ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs and Domains	Homologous Sequences	Analytical Methods and Databases
53	880	S18 S68 T123 T143 S159 T178 T286 S294 S327 S376 S388 T397 T403 S426 S438 S474 S563 T587 T634 T645 S659 S665 S677 S756 S799 S809 T827 S870 S82 T88 S99 T131 T165 S215 S253 S362 S487 T510 S525 S589 T593 S622	N60 N251 N338 N514 N585 N643	Myb1 DNA-binding domain: W808-I816 WD40 domains: L41-N79, K84-N124, T131-D170, G239-D281, A771-S809, F157-T171 Acidic Serine Cluster Repeat: A423-R697	homologous to mouse gene PC326 g458692 (Homo sapiens) Bergsagel, P.L. et al. (1992) Oncogene 7:2059-2064.	BLAST-GenBank BLAST-DOMO HMMER-PFAM BLIMPS-PRINTS MOTIFS
54	855	T460 S8 S179 S261 T288 T313 T377 T706 T719 T755 S764 S803 S851 S34 S67 T129 S190 S339 T391 S483 S502 S537 Y92	N552	Crooked neck protein (RNA processing associated, contains TPR repeat): W398-V814	Predicted TPR domain protein G2315362 (Caenorhabditis elegans) Zhang, K. et al. (1991) Genes Dev. 5:1080-1091.	BLAST-GenBank BLAST-PRODOM MOTIFS

Table 3

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
55	263-307	Cardiovascular (0.200) Gastrointestinal (0.200) Reproductive (0.200)	Cancer (0.433) Inflammation (0.267) Cell Proliferation (0.200)	PBLUESCRIPT
56	406-450	Reproductive (0.222) Cardiovascular (0.167) Gastrointestinal (0.167) Nervous (0.167)	Cancer (0.500) Inflammation (0.389) Cell Proliferation (0.167)	PSPORT1
57	1001-1045	Reproductive (0.265) Gastrointestinal (0.206) Nervous (0.206)	Cancer (0.412) Inflammation (0.324) Cell Proliferation (0.176)	PINCY
58	226-270	Nervous (0.316) Hematopoietic/Immune (0.211) Reproductive (0.211)	Cancer (0.368) Inflammation (0.368) Cell Proliferation (0.158)	PINCY
59	406-450	Hematopoietic/Immune (0.500) Cardiovascular (0.227)	Cancer (0.182) Inflammation (0.682) Cell Proliferation (0.136)	PINCY
60	56-100	Gastrointestinal (0.545) Nervous (0.182) Reproductive (0.182)	Cancer (0.545) Inflammation (0.364) Cell Proliferation (0.273)	PINCY
61	1046-1090	Nervous (0.271) Reproductive (0.220) Gastrointestinal (0.153)	Cancer (0.542) Inflammation (0.288) Cell Proliferation (0.220)	PINCY
62	226-270	Hematopoietic/Immune (0.288) Nervous (0.178) Reproductive (0.164)	Cancer (0.397) Inflammation (0.548)	PINCY
63	559-603	Reproductive (0.260) Gastrointestinal (0.145) Cardiovascular (0.130)	Cancer (0.458) Inflammation (0.359) Cell Proliferation (0.176)	PSPORT1
64	12-56	Reproductive (0.385) Gastrointestinal (0.231) Cardiovascular (0.154) Nervous (0.154)	Cancer (0.538) Inflammation (0.154) Cell Proliferation (0.154)	PINCY
65	488-532 1091-1135	Reproductive (0.308) Nervous (0.282) Gastrointestinal (0.154)	Cancer (0.487) Inflammation (0.231) Cell Proliferation (0.103)	PINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
66	37-81	Nervous (0.500) Dermatologic (0.250) Reproductive (0.250)	Inflammation (0.500)	PINCY
67	326-370 1136-1180	Nervous (0.237) Reproductive (0.237) Hematopoietic/Immune (0.158)	Cancer (0.395) Inflammation (0.316) Cell Proliferation (0.158)	PINCY
68	451-495	Nervous (0.312) Reproductive (0.312) Developmental (0.125) Hematopoietic/Immune (0.125) Urologic (0.125)	Cancer (0.562) Inflammation (0.188) Cell Proliferation (0.312)	PINCY
69	64-108	Reproductive (0.233) Nervous (0.174) Cardiovascular (0.140)	Cancer (0.477) Inflammation (0.279) Cell Proliferation (0.198)	PINCY
70	77-121	Cardiovascular (0.500) Musculoskeletal (0.500)	Cancer (0.500) Trauma (0.500)	PBLUESCRIPT
71	164-208	Developmental (0.222) Nervous (0.222)	Cancer (0.444) Cell proliferation (0.222) Trauma (0.222)	PINCY
72	604-648	Reproductive (0.362) Gastrointestinal (0.149) Hematopoietic/Immune (0.128)	Cancer (0.426) Inflammation/Trauma (0.276) Cell proliferation (0.170)	PINCY
73	106-150 1066-1110	Reproductive (0.307) Nervous (0.202) Cardiovascular (0.114)	Cancer (0.482) Inflammation/Trauma (0.307) Cell proliferation (0.175)	PINCY
74	651-695	Hematopoietic/Immune (0.290) Reproductive (0.226) Cardiovascular (0.161)	Inflammation/Trauma (0.451) Cell proliferation (0.230) Cancer (0.320)	PINCY
75	241-285 535-579	Reproductive (0.193) Cardiovascular (0.169) Gastrointestinal (0.157)	Cancer (0.458) Inflammation/Trauma (0.337) Cell proliferation (0.169)	PINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
76	173-217 593-637	Nervous (0.513) Reproductive (0.167)	Inflammation/Trauma (0.371) Cancer (0.333) Cell proliferation (0.141)	PINCY
77	13-57	Reproductive (0.241) Nervous (0.202) Cardiovascular (0.140)	Cancer (0.461) Inflammation (0.180) Cell Proliferation (0.167)	PBLUESCRIPT
78	176-220	Nervous (0.279) Reproductive (0.235) Gastrointestinal (0.147)	Cancer (0.500) Inflammation (0.176) Cell Proliferation (0.162)	PBLUESCRIPT
79	79-123	Nervous (0.280) Cardiovascular (0.160) Developmental (0.160)	Cancer (0.480) Cell Proliferation (0.480) Inflammation (0.160)	PBLUESCRIPT
80	870-914	Nervous (0.571) Reproductive (0.238) Developmental (0.095)	Cancer (0.238) Inflammation (0.381) Cell Proliferation (0.190)	PSPORT1
81	149-194	Nervous (0.216) Reproductive (0.201) Gastrointestinal (0.185)	Cancer (0.432) Inflammation (0.259) Cell Proliferation (0.154)	PINCY
82	150-194	Reproductive (0.375) Cardiovascular (0.125) Endocrine (0.125) Hematopoietic/Immune (0.125) Developmental (0.125) Urologic (0.125)	Cancer (0.375) Inflammation (0.375) Trauma (0.250)	PSPORT1
83	177-221	Reproductive (0.199) Gastrointestinal (0.173) Hematopoietic/Immune (0.128) Nervous (0.128)	Cancer (0.429) Inflammation (0.270) Cell Proliferation (0.186)	PINCY
84	342-386	Reproductive (0.252) Gastrointestinal (0.196) Nervous (0.161)	Cancer (0.483) Inflammation (0.238) Cell Proliferation (0.161)	PINCY
85	124-168	Hematopoietic/Immune (0.308) Cardiovascular (0.154) Nervous (0.154) Gastrointestinal (0.154)	Cancer (0.538) Inflammation (0.308)	PINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
86	238-282	Reproductive (0.277) Cardiovascular (0.181) Nervous (0.169)	Cancer (0.434) Inflammation (0.193) Cell Proliferation (0.157)	PINCY
87	117-161	Reproductive (0.250) Gastrointestinal (0.250) Hematopoietic/Immune (0.115)	Cancer (0.558) Inflammation (0.192) Cell Proliferation (0.115) Trauma (0.115)	PINCY
88	139-183	Nervous (0.237) Reproductive (0.214) Gastrointestinal (0.168)	Cancer (0.397) Inflammation (0.298) Trauma (0.137)	PINCY
89	184-228 352-396	Reproductive (0.556) Nervous (0.222) Hematopoietic/Immune (0.111) Developmental (0.111)	Cancer (0.444) Inflammation (0.333) Cell Proliferation (0.333)	PINCY
90	69-113 879-923	Nervous (0.316) Reproductive (0.193) Hematopoietic/Immune (0.158)	Cancer (0.439) Inflammation (0.211) Cell Proliferation (0.123)	PINCY
91	72-116	Nervous (0.211) Reproductive (0.197) Gastrointestinal (0.158)	Cancer (0.461) Inflammation (0.263) Cell Proliferation (0.211)	PSPORT1
92	489-533	Reproductive (0.274) Nervous (0.217) Gastrointestinal (0.123)	Cancer (0.481) Inflammation (0.189) Cell Proliferation (0.160)	PSPORT1
93	761-805	Reproductive (0.219) Hematopoietic/Immune (0.156) Developmental (0.125)	Cancer (0.312) Cell Proliferation (0.281) Inflammation (0.188) Trauma (0.188)	PSPORT1
94	126-170	Reproductive (0.379) Nervous (0.241) Developmental (0.138)	Cancer (0.414) Cell Proliferation (0.241) Inflammation (0.103)	PBLUESCRIPT
95	1173-1217	Reproductive (0.192) Gastrointestinal (0.192) Nervous (0.173)	Cancer (0.481) Inflammation (0.250) Cell Proliferation (0.212)	PINCY
96	465-509	Hematopoietic/Immune (0.250) Cardiovascular (0.158) Gastrointestinal (0.145)	Inflammation (0.368) Cancer (0.355) Cell Proliferation (0.132)	PINCY

Table 3 (cont.)

Nucleotide Seq ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
97	2427-2471	Nervous (0.224) Reproductive (0.197) Gastrointestinal (0.184)	Cancer (0.474) Cell Proliferation(0.263) Inflammation (0.237)	PINCY
98	23-67	Gastrointestinal (0.270) Reproductive (0.190) Cardiovascular (0.135)	Cancer (0.429) Inflammation (0.278) Cell Proliferation(0.143)	PINCY
99	106-150	Gastrointestinal (0.263) Reproductive (0.263) Nervous (0.158)	Cancer (0.474) Inflammation (0.368) Cell Proliferation(0.211)	PINCY
100	73-117 460-504	Hematopoietic/Immune (0.211) Reproductive (0.211) Cardiovascular (0.105) Developmental (0.105) Gastrointestinal (0.105) Musculoskeletal (0.105)	Cancer (0.474) Inflammation (0.263) Cell Proliferation(0.211)	PSPORT1
101	861-905	Developmental (0.333) Nervous (0.667)	Cell Proliferation(0.333) Trauma (0.333) Neurological (0.333)	PINCY
102	8-52	Developmental (1.000)	Cell Proliferation (1.000)	PINCY
103	199-243	Hematopoietic/Immune (0.143) Nervous (0.179) Reproductive (0.286)	Cancer (0.536) Inflammation (0.250) Cell Proliferation(0.214)	PINCY
104	413-457 908-952	Nervous (0.236) Reproductive (0.222) Gastrointestinal (0.125)	Cancer (0.458) Inflammation (0.236) Cell Proliferation(0.139)	PINCY
105		Reproductive (0.270) Gastrointestinal (0.169) Hematopoietic/Immune 0.101) Developmental (0.101) Nervous (0.101)	Cancer (0.449) Inflammation (0.281) Cell Proliferation(0.258)	PINCY
106	255-299 513-557	Reproductive (0.216) Gastrointestinal (0.196) Nervous (0.157)	Cancer (0.490) Inflammation (0.176) Cell Proliferation(0.176)	PINCY
107	167-211 814-859 1922-1966	Reproductive (0.263) Nervous (0.162) Gastrointestinal (0.141)	Cancer (0.455) Inflammation (0.202) Trauma (0.131)	PINCY



Table 3 (cont.)

Nucleotide Seq ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition Fraction of Total	Vector
108	877-921 2230-2274	Reproductive (0.299) Nervous (0.206) Gastrointestinal (0.134)	Cancer (0.536) Inflammation (0.227) Cell Proliferation(0.124)	PINCY

WO 01/07471

PCT/US00/19948

Table 4

Nucleotide SEQ ID NO:	Library	Library Description
55	KIDNNOT01	Library was constructed using RNA isolated from the kidney tissue of a 64-year-old Caucasian female, who died from an intracranial bleed. Patient history included rheumatoid arthritis.
56	BRSTNOT02	Library was constructed using RNA isolated from diseased breast tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated proliferative fibrocystic changes characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Pathology for the associated tumor tissue indicated an invasive grade 4 mammary adenocarcinoma. Patient history included atrial tachycardia and a benign neoplasm. Family history included cardiovascular and cerebrovascular disease.
57	PLACNOT02	Library was constructed using RNA isolated from the placental tissue of a Hispanic female fetus, who was prematurely delivered at 21 weeks' gestation. Serologies of the mother's blood were positive for CMV (cytomegalovirus).
58	BRAINOT12	Library was constructed using RNA isolated from brain tissue removed from the right frontal lobe of a 5-year-old Caucasian male during a hemispherectomy. Pathology indicated extensive polymicrogyria and mild to moderate gliosis (predominantly subpial and subcortical), which are consistent with chronic seizure disorder. Family history included a cervical neoplasm.
59	SPLNNOT04	Library was constructed using RNA isolated from the spleen tissue of a 2-year-old Hispanic male, who died from cerebral anoxia.
60	LNDNOT03	Library was constructed using RNA isolated from lymph node tissue obtained from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy. On microscopic exam, this tissue was found to be extensively necrotic with 10% viable tumor. Pathology for the associated tumor tissue indicated invasive grade 3-4 squamous cell carcinoma. Patient history included hemangioma. Family history included atherosclerotic coronary artery disease, benign hypertension, congestive heart failure, atherosclerotic coronary artery disease.
61	LIVRTUT01	Library was constructed using RNA isolated from liver tumor tissue removed from a 51-year-old Caucasian female during a hepatic lobectomy. Pathology indicated metastatic grade 3 adenocarcinoma consistent with colon cancer. Family history included a malignant neoplasm of the liver.

Table 4 (cont.)

Nucleotide Seq ID NO:	Library	Library Description
62	BLADPUT07	Library was constructed using RNA isolated from bladder tumor tissue removed from the anterior bladder wall of a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrostomy. Pathology indicated a grade 3 transitional cell carcinoma in the left lateral bladder. Patient history included angina, emphysema, and tobacco use. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
63	LUNGAST01	Library was constructed using RNA isolated from the lung tissue of a 17-year-old Caucasian male, who died from head trauma. Patient history included asthma.
64	LIVREF02	Library was constructed using RNA isolated from liver tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
65	LUNGNOT23	Library was constructed using RNA isolated from left lobe lung tissue removed from a 58-year-old Caucasian male. Pathology for the associated tumor tissue indicated metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Family history included prostate cancer, breast cancer, and acute leukemia.
66	TESTNOT07	Library was constructed using RNA isolated from testicular tissue removed from a 31-year-old Caucasian male during an unilateral orchiectomy (excision of testis). Pathology indicated a mass containing a large subcapsular hematoma with laceration of the tunica albuginea. The surrounding testicular parenchyma was extensively necrotic.
67	PROSTUT13	Library was constructed using RNA isolated from prostate tumor tissue removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3). Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis. Family history included multiple myeloma, hyperlipidemia, and rheumatoid arthritis.
68	LNODNOT11	Library was constructed using RNA isolated from lymph node tissue removed from a 16-month-old Caucasian male who died from head trauma. Patient history included bronchitis.

101

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
69	BRSTNOT35	Library was constructed using RNA isolated from breast tissue removed from a 46-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated normal breast parenchyma, bilaterally. The patient presented with hypertrophy of breast and headache. Patient history included obesity, lumbago, glaucoma, and alcohol abuse. Family history included cataract, osteoarthritis, uterine cancer, benign hypertension, hyperlipidemia, alcoholic cirrhosis of the liver, cerebrovascular disease, and type II diabetes.
70	MUSCNOT01	Library was constructed at Stratagene (STR937209), using RNA isolated from the skeletal muscle tissue of a patient with malignant hyperthermia.
71	LUNGNOT14	Library was constructed using RNA isolated from lung tissue removed from the left lower lobe of a 47-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated a grade 4 adenocarcinoma, and the parenchyma showed calcified granuloma. Patient history included benign hypertension and chronic obstructive pulmonary disease. Family history included type II diabetes and acute myocardial infarction.
72	UTRSNOT06	Library was constructed using RNA isolated from myometrial tissue removed from a 50-year-old Caucasian female during a vaginal hysterectomy. Pathology indicated residual atypical complex endometrial hyperplasia. Pathology for the associated tissue removed during dilation and curettage indicated fragments of atypical complex hyperplasia and a single microscopic focus suspicious for grade 1 adenocarcinoma. Patient history included benign breast neoplasm, hypothyroid disease, polypectomy, and arthralgia. Family history included cerebrovascular disease, atherosclerotic coronary artery disease, hyperlipidemia, and chronic hepatitis.
73	PROSTUT08	Library was constructed using RNA isolated from prostate tumor tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated an adenocarcinoma (Gleason grade 3+4). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst, and hematuria. Family history included tuberculosis, cerebrovascular disease, and arteriosclerotic coronary artery disease.
74	THYMNOT03	Library was constructed using RNA isolated from thymus tissue removed from a 21-year-old Caucasian male during a thymectomy. Pathology indicated an unremarkable thymus and a benign parathyroid adenoma in the right inferior parathyroid. Patient history included atopic dermatitis, a benign neoplasm of the parathyroid, and tobacco use. Family history included atherosclerotic coronary artery disease and benign hypertension.

Table 4 (cont.)

Nucleotide Seq ID NO:	Library	Library Description
75	PENCNOT01	Library was constructed using RNA isolated from penis corpus cavernosum tissue removed from a 53-year-old male. Patient history included untreated penile carcinoma.
76	BRAUNOT01	Library was constructed using RNA isolated from caudate/putamen/nucleus accumbens tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomenigeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.
77	HUVELPB01	This library was constructed using RNA isolated from HUV-EC-C (ATCC CRL 1730) cells stimulated with cytokine/LPS. RNA was isolated from two pools of HUV-EC-C cells that had been treated with either 4 units/ml TNF-alpha and 2 units/ml gamma IFN for 96 hours, or 1 unit/ml IL-1 beta and 100 ng/ml LPS for 5 hours.
78	HUVENOB01	This library was constructed using RNA isolated from HUV-EC-C (ATCC CRL 1730) cells.
79	HMT2RAT01	This library was constructed at Stratagene (STR937231), using RNA isolated from the hMT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated with retinoic acid for 24 hours.
80	BRAINOT04	This library was constructed using RNA isolated from the brain tissue of a 44-year-old Caucasian male with a cerebral hemorrhage. The tissue, which contained coagulated blood, came from the choroid plexus of the right anterior temporal lobe. Family history included coronary artery disease and myocardial infarction.
81	BRAITUT08	This library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 47-year-old Caucasian male during excision of cerebral meningeal tissue. Pathology indicated grade 4 fibrillary astrocytoma with focal tumoral radionecrosis. Patient history included cerebrovascular disease, deficiency anemia, hyperlipidemia and epilepsy. Family history included cerebrovascular disease and a malignant prostate neoplasm.
82	PROSNON01	This library was constructed from 4.4 million independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
83	PANCTUT01	This library was constructed using RNA isolated from pancreatic tumor tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Previous surgeries included a total splenectomy, cholecystectomy, and abdominal hysterectomy. Family history included cardiovascular disease, type II diabetes, and stomach cancer.
84	BRAITUT13	This library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 68-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a meningioma in the left frontal lobe.
85	STOMFET01	This library was constructed using RNA isolated from the stomach tissue of a Caucasian female fetus, who died at 20 weeks' gestation.
86	PROSNOT16	This library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
87	SINTNOT13	This library was constructed using RNA isolated from ileum tissue obtained from a 25-year-old Asian female during a partial colectomy and temporary ileostomy. Pathology indicated moderately active chronic ulcerative colitis, involving colonic mucosa from the distal margin to the ascending colon. Family history included hyperlipidemia, depressive disorder, malignant cervical neoplasm, viral hepatitis A, and depressive disorder.
88	SINTNOT13	This library was constructed using RNA isolated from ileum tissue obtained from a 25-year-old Asian female during a partial colectomy and temporary ileostomy. Pathology indicated moderately active chronic ulcerative colitis, involving colonic mucosa from the distal margin to the ascending colon. Family history included hyperlipidemia, depressive disorder, malignant cervical neoplasm, viral hepatitis A, and depressive disorder.
89	LUNGFET03	This library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
90	SKINBIT01	This library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.

Table 4 (cont.)

Nucleotide Seq ID NO:	Library	Library Description
91	LUNGCTUT03	This library was constructed using RNA isolated from lung tumor tissue removed from the left lower lobe of a 69-year-old Caucasian male during segmental lung resection. Pathology indicated residual grade 3 invasive squamous cell carcinoma. Patient history included acute myocardial infarction, prostatic hyperplasia, malignant skin neoplasm, and tobacco use.
92	OVARUTUT01	This library was constructed using RNA isolated from ovarian tumor tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology indicated grade 2 mucinous cystadenocarcinoma involving the entire left ovary. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.
93	LUNGFET05	This library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation from anencephalus.
94	ENDANOT01	This library was constructed using RNA isolated from aortic endothelial cell tissue from an explanted heart removed from a male during a heart transplant.
95	ESOGTUT02	This library was constructed using RNA isolated from esophageal tumor tissue obtained from a 61-year-old Caucasian male during a partial esophagectomy, proximal gastrectomy, pyloromyotomy, and regional lymph node excision. Pathology indicated an invasive grade 3 adenocarcinoma in the esophagus. Family history included atherosclerotic coronary artery disease, type II diabetes, chronic liver disease, primary cardiomyopathy, benign hypertension, and cerebrovascular disease.
96	SINIUCT01	This library was constructed using RNA isolated from ileum tissue obtained from a 42-year-old Caucasian male during a total intra-abdominal colectomy and endoscopic jejunostomy. Previous surgeries included polypectomy, colonoscopy, and spinal canal exploration. Family history included cerebrovascular disease, benign hypertension, atherosclerotic coronary artery disease, and type II diabetes.
97	NPOLNOT01	This library was constructed using RNA isolated from nasal polyp tissue removed from a 78-year-old Caucasian male during a nasal polypectomy. Pathology indicated a nasal polyp and striking eosinophilia. Patient history included asthma and nasal polyps.
98	ADRENOT09	This library was constructed using RNA isolated from left adrenal gland tissue removed from a 43-year-old Caucasian male during nephroureterectomy, regional lymph node excision, and unilateral left adrenalectomy. Pathology for the associated tumor tissue indicated a grade 2 renal cell carcinoma mass in the posterior lower pole of the left kidney with invasion into the renal pelvis.

Table 4 (cont.)

Nucleotide Seq ID NO:	Library	Library Description
99	BRAIUNT01	This library was constructed using RNA isolated from SK-N-MC, a neuroepithelioma cell line (ATCC HTB-10) derived from a 14-year-old Caucasian female with neuroepithelioma, with metastasis to the supra-orbital area.
100	LUNGNON03	This library was constructed from 2.56 x 1e6 independent clones from a lung tissue library. RNA was made from lung tissue removed from the left lobe a 58-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated a metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Patient also received radiation therapy to the retroperitoneum. Family history included prostate cancer, breast cancer, and acute leukemia. The normalization and hybridization conditions were adapted from Soares et al., PNAS (1994) 91:9228; Swaroop et al., NAR (1991) 19:1954; and Bonaldo et al., Genome Research (1996) 6:791.
101	BRADDIT02	This library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, and emphysema.
102	PLACNOT05	This library was constructed using RNA isolated from placental tissue removed from a Caucasian male fetus, who died after 18 weeks' gestation from fetal demise.
103	HELATXT03	This library was constructed using RNA isolated from a treated HeLa cell line, derived from cervical adenocarcinoma removed from a 31-year-old Black female. The cells were treated with 1 microm PMA and 100 microm cycloheximide for 24 hours.
104	COLHTUT01	This library was constructed using RNA isolated from colon tumor tissue removed from the hepatic flexure of a 55-year-old Caucasian male during right hemicolectomy, incidental appendectomy, and permanent colostomy. Pathology indicated invasive grade 3 adenocarcinoma. Patient history included benign hypertension, anxiety, abnormal blood chemistry, blepharitis, heart block, osteoporosis, acne, and hyperplasia of prostate. Family history included prostate cancer, acute myocardial infarction, stroke, and atherosclerotic coronary artery disease.
105	PLACFER01	This library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus who died after 16 weeks' gestation from fetal demise and hydrocephalus. Serology was positive for CMV antibody.
106	293TF2T01	This library was constructed using RNA isolated from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. The cells were treated with 5-aza-2'-deoxycytidine and transformed with adenovirus 5 DNA.



Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
107	BRAENOT02	This library was constructed using RNA isolated from posterior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male.
108	FTUBTUT02	This library was constructed using RNA isolated from fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid and serous adenocarcinoma confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma in an endometrial polyp. Metastatic endometrioid and serous adenocarcinoma were present. The patient presented with a pelvic mass and ascites. Patient history included medullary carcinoma of the thyroid and myocardial infarction.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	PE Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	PE Biosystems, Foster City, CA; Parcel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	PE Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, ifasta, fastx, ifastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A Blocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; S. nhhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group

consisting of:

- 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54,

- 15 b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54,

25 c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54,

30 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54, and

35 of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54, and

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfilesScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gritskov, M. et al. (1988) CABIOS 4:61-66; Gritskov, M. et al. (1989) Methods Enzymol. 183: 146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score>GCC-specified "HGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8: 186-194.	
Phrap	A Phis Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

5 NO:15, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, and SEQ ID NO:108.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method for producing a polypeptide of claim 1, the method comprising:
- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, and SEQ ID NO:108,

- b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, and SEQ ID NO:108,

- c) a polynucleotide sequence complementary to a),
- d) a polynucleotide sequence complementary to b), and
- e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4,

SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, and SEQ ID NO:54.

18. A method for treating a disease or condition associated with decreased expression of

functional CCYPR, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional CCYPR, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional CCYPR, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:



a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.



(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
1 February 2001 (01.02.2001)

PCT

(10) International Publication Number  
WO 01/07471 A3

(51) International Patent Classification:  
C12N 15/12, 5/10, C07K 14/47, 16/18, C12Q 1/68, A61K 38/17, G01N 33/50, A01K 67/027

(21) International Application Number: PCT/US00/19948  
(22) International Filing Date: 21 July 2000 (21.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/145,075  
21 July 1999 (21.07.1999) US  
60/153,129  
8 September 1999 (08.09.1999) US  
60/164,647  
10 November 1999 (10.11.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

60/145,075 (CIP)  
US  
Filed on 21 July 1999 (21.07.1999)  
60/153,129 (CIP)  
US  
Filed on 8 September 1999 (08.09.1999)  
60/164,647 (CIP)  
US  
Filed on 10 November 1999 (10.11.1999)

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and  
(75) Inventors/Applicants (for US only): HILTMAN, Jennifer, L. [US/US]; 230 Monroe Drive #17, Mountain View, CA 94040 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). AZIMZAI, Yalda [US/US]; 5518 Boulder

Canyon Drive, Castro Valley, CA 94552 (US). YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). BAUGHN, Mariah, R. [US/US]; 14244 Sautago Road, San Leandro, CA 94577 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). SHAH, Purvi [IN/US]; 1608 Queen Charlotte Drive #5, Sunnyvale, CA 94087 (US).  
(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, GR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published: — with international search report

(88) Date of publication of the international search report: 17 January 2002

(15) Information about Correction:

Previous Correction: see PCT Gazette No. 20/2001 of 17 May 2001, Section II  
For two-letter codes and other abbreviations, refer to the "Guide to the Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CELL CYCLE AND PROLIFERATION PROTEINS

(57) Abstract: The invention provides human cell cycle and proliferation proteins (CCYPR) and polynucleotides which identify and encode CCYPR. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of CCYPR.

WO 01/07471 A3



# INTERNATIONAL SEARCH REPORT

Inte al Application No

PCT/US 00/19948

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C12N5/10 C07K14/47 C07K16/18 C12Q1/68  
A61K38/17 G01N33/50 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 C12N C07K C12Q A61K G01N A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, where appropriate, of the relevant passages	Relevant to claim No.
-----------	---	-----------------------

X	NEMOTO Y ET AL: "Recruitment of an alternatively spliced form of synaptotagmin 2 to mitochondria by the interaction with the PDZ domain of a mitochondrial outer membrane protein" EMBO JOURNAL, vol. 18, no. 11, 1 June 1999 (1999-06-01), pages 2991-3006, XP002156389 Rat OMP25: 88.966% identity in 145 aa overlap with SeqIdNo.1 / 75.835% identity in 1167 nt overlap with SeqIdNo.55 MO 98 45436 A (GENETICS INST) 15 October 1998 (1998-10-15) SeqIdNo.1414: 99.8% identity in 432 bp overlap with SeqIdNo.55 --- -/-	3,11,12 1,3,6,7, 9-11,13, 15,19, 22,25,26
---	--	---

<input checked="" type="checkbox"/>	Further documents are listed in the continuation of box C.	<input checked="" type="checkbox"/> Patent family members are listed in annex.
-------------------------------------	--	--

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other documents, such combination being obvious to a person skilled in the art
- "Z" document member of the same patent family

Date of the actual completion of the international search	4 January 2001
Date of mailing of the international search report	25 04 2001

Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Lonnoy, O Authorized officer
--	---------------------------------

4

5

6

7

# INTERNATIONAL SEARCH REPORT

Inte  
Application No  
PCT/US 00/19948

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
----------	--	-----------------------

E	EP 1 033 401 A (GENSET) 6 September 2000 (2000-09-06) SeqIdNo.3623: 100.000% identity in 374 nt overlap with SeqIdNo.55 -& DATABASE GENESSEQ [online] E.B.I., Hinxton, U.K.; Accession Number: C03625, 6 October 2000 (2000-10-06) DUMAS M ET AL: "Human secreted protein 5" EST, SeqIdNo.3623" XP002156390 abstract	1,3,6,7, 9-15
T		
A	WO 97 12962 A (COLD SPRING HARBOR LAB :BEACH DAVID (US); CALIGIURI MAUREEN (US);) 10 April 1997 (1997-04-10)	





INTERNATIONAL SEARCH REPORT

B x I Observations where certain claims were found unsatisfactory (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

B x II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see further information sheet invention group 1.

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.



FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claim : 1



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No  
PCT/US 00/19948

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
---	---------------------	----------------------------	---------------------

WO 9845436	A	AU 6891098 A	30-10-1998
EP 1033401	A	EP 0973896 A	26-01-2000
EP 1033401	A	NONE	
WO 9712962	A	US 6001619 A	14-12-1999
		EP 0857205 A	12-08-1998

